

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 05:00:24 ; Search time 2731 Seconds
(without alignments)
3666.585 Million cell updates/sec

Title: US-09-889-756A-2
Perfect score: 2019
Sequence: 1 MAFYAFMRALAAALAAVAL.....AAPOSGVOTASEAKTASAE 412

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09889756/runat.05092003.124325.23309/app.query.fasta_1.583
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09889756 @CGN 1.1 2810 @runat.05092003.124325.23309 -NCPU=6 -ICPU=3
-NO MAP -LARGEOUTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vit:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	585	29.0	893	29	CNS01JMC
2	512.5	25.4	834	29	CNS01JMC
3	411.5	20.4	642	28	BH190105
4	370.5	18.4	601	28	BH190105
5	343	17.0	1096	29	BZ558375
6	318.5	15.8	1503	29	BZ569193
7	315	15.6	775	29	BZ550352
8	304	15.1	515	28	BH190022
9	291.5	14.4	1236	29	BZ554554
10	280	13.9	798	28	BH250378
11	269.5	13.3	620	28	BH189811
12	261.5	13.0	1263	29	BZ557583
13	254	12.6	937	29	BZ562652
14	250	12.4	1417	29	BZ574393
15	249	12.3	802	10	BF663822
16	246	12.2	620	28	BH190195
17	244	12.1	596	28	BH189826
18	241.5	12.0	1108	29	BZ577818
19	233	11.5	1259	29	BZ580098
20	220	10.9	470	29	BZ332913
21	219.5	10.9	923	29	BZ554963
22	216	10.7	597	28	BH189931
23	216	10.7	747	29	BZ561644
24	211.5	10.5	568	29	BZ336018
25	210.5	10.4	772	29	BZ579248
26	194	9.6	809	29	BZ551287
27	189.5	9.4	595	14	CA554488
28	189.5	9.4	766	14	BZ558130
29	188	9.3	775	28	AZ933876
30	186	9.2	607	29	BZ340887
31	186	9.2	1379	29	BZ569194
32	180.5	8.9	472	28	AZ302989
33	179.5	8.9	450	28	BH247218
34	177	8.8	745	28	BH242325
35	173.5	8.6	847	29	BZ562720
36	172.5	8.5	793	29	BZ549534
37	170.5	8.4	876	29	BZ638338
38	168.5	8.3	469	10	BF416737
39	168	8.3	1307	29	BZ574481
40	163.5	8.1	1228	29	BZ578763
41	163	8.1	811	29	BZ577024
42	161	8.0	867	29	BZ566980
43	160	7.9	1191	25	BZ552701
44	159	7.9	615	29	BZ345306
45	158.5	7.9	1083	29	BZ579554

ALIGNMENTS

RESULT 1
CNS01JMC 893 bp DNA linear GSS 12-JUN-2001
LOCUS Anopheles gambiae GSS T7 end of clone 13008 of NotreDame1 library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL147221
VERSION AL147221.1 GI:7005367
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae (African malaria mosquito)
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

REFERENCE	1 (bases 1 to 893)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
REFERENCE	2 (bases 1 to 893)
AUTHORS	Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE	Direct Submission
JOURNAL	Submitted (16-FEB-2000) BIMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT	This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES	Location/Qualifiers
source	1..893
	/organism="Anopheles gambiae"
	/mol_type="genomic DNA"
	/strain="PEST"
	/db_xref="taxon:7165"
	/clone="13008"
	/clone_lib="Notredame1"
	/note="Tend : 77"
BASE COUNT	174 a 247 c 246 g 221 t 5 others
ORIGIN	
Alignment Scores:	
Pred. No.:	3,34e-51
Score:	585.00
Percent Similarity:	64.67%
Best Local Similarity:	42.67%
Query Match:	28.97%
DB:	29 Gaps: 2
US-09-889-756A-2 (1-412) x CNS01JMC (1-893)	
OY	85 LysATgLeuPheGInGluGlySerTyValArgAlaGlyGlnProLeuTyrgInileasp 104
Db	888 AAACGCACTTACCGAAGCGGTATGTGCAGSCCGGTGAGCTCTGATCAGATTGAT 829
OY	105 SerSerThrTyGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db	828 CCGCGAACCTTACAGCGCTTATGAAACGGGAAAGGGAGATCTGGCTAAAGCACGGCC 769
OY	125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeuValAlaAlaGlnAla 144
Db	768 GCGCGTAAATATGCCCC-CTGACGCTGAACCGCTATCAAAAACCTGCTCGTACCAATAC 710
OY	145 ValSerArgGInGluTyArgAspAlaAlaValAlaThrAlaLysArgSerAlaGlnAlaVal 164
Db	709 ATCAGTCAGCAGATTACGATACCGGCTCGGGGAGATGCCCCAGCCAGTAACGCCGCGTg 650
OY	165 LysAlaAlaGlnAlaAlaLysSerAlaGlyLeuAsnLeuAsnArgSerArgIleThr 184
Db	649 GTGGCAGCCAAAGCCGCGCTCGAAACCGCGGATTAACCTGGCTTATACCAACGTGACC 590
OY	185 AlaProlIleSerGlyPheIleGlyGInSerLysValSerGluGlyThrLeuLeuAsnAla 204
Db	589 TCCTCTATCAGCGCTCGTATGTGTAATCTCTCCGTGACGGAAGGGCTCTGTGTCCGAAC 530
OY	205 GlyAspThrThrValLeuAlaThrIleArgGInThrAsnProMetTyValAsnValThr 224
Db	529 GGTCAAGNACATCGCGCTGGCAACCGTGCAGAGTGCATCCGATCTATGTTCAGTCAAG 470
OY	225 GInSerAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLysLeuAla 244
Db	469 CAGTCAGCAATGATTTCTCGCGCTCGAAACAGAGACTGGCTAACCGCACCTGAAACAG 410
OY	245 AlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProGluLys 264
Db	409 GAAAACGGGAAAGCCAAAGTGAAGCTGATTACCAAGACGATATCAAGTTCCCGCAGGAA 350

OY		265	G yArtrgLeuNeUPheAlAAsPProValASndIuseerThrGInglIEthLEuArg	284
Dd		349	GGAGCGCTGAATTCTGTACGTACCGTACCAGAACCCGGTTTCATCCTTAGCT	290
OY		285	A Aa Va P oASnaePg nASnleUmeWPrGtlyLeuTyVAlArGValLauMet	304
Dd		289	GCAGTTTTCCCAGAACCTTGACAAMAAATCTGCTCCAGSTAGTGTGTCGCCAGCTGTG	230
OY		305	ASpglnVA AlaValASPasnAlAPheValVProlnglncinalAVAlThrArg-----	322
Dd		229	GAAAGAAGAACGAAATCAACCGCCCTTGTGTTCACAGCAGGGGTGTGACCCGTASCAC	170
OY		323	G y ALysAsPTrrValMeTlleValASnlaglInclyglyMeGluPProArgGlVal	342
Dd		169	GC GGCGATGCG-AACGCACGTGTTGTTGGCGCTATAACAAGTCGAATCCGAATATC	111
OY		343	ThrValAagIngIngInglyThrAsnTrpIleValThrserylEuleuYasPGlyAsp	362
Dd		110	ACCGCACCCACGGCATTTGGGGATTAATGGCTGTGACGMAAGGTCTGAAGATGCGCAT	51
OY		363	LysValValValGluGlyIleSerIleAlaGlyLehthrGlyAlaybLYsValThrPro	382
Dd		50	CGCGTGAATTGT-----ACTGGTTTGCAAAAAGTTCGTCT	15
RESULT 2				
CNSO1JfPX/c				
LOCUS		CNSO1JPX	834 bp DNA linear GSS J2-JUN-2001	
DEFINITION		Anopheles gambiae GSS T7 end of clone 14B21 of NotreDame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.		
ACCESSION		AF147350		
KEYWORDS		ALI47350.1 GI:7005496		
SOURCE		GSS.		
ORGANISM		Anopheles gambiae (African malaria mosquito)		
Eukaryota; Metazoa;				
Insecta; Diptera;				
Nematocera; Culicoidae;				
Anophelinae.				
REFERENCE		1 (bases 1 to 834)		
AUTHORS		Genoscope.		
TITLE		Direct Submission		
JOURNAL		Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY Cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
2 (bases 1 to 834)				
REFERENCE		Roch,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.		
AUTHORS		Direct Submission		
TITLE		Submitted (16-FEB-2000) BMWI, Institut Pasteur, 25, rue du Dr.		
JOURNAL		Roux, Paris 75015, France		
COMMENT		This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. Of Insects, Institut Pasteur.		
FEATURES				
source		Location/Qualifiers		
		1..834		
		/organism="Anopheles gambiae"		
		/mol_type="Genomic DNA"		
		/strain="PEST"		
		/db_xref="taxon:7165"		
		/clone="14B21"		
		/clone_id="NotreDame1"		
		/note="end : T7"		
BASE COUNT		156 a 227 c 229 g 214 t	8 others	
ORIGIN				
Alignment Scores:				
Pred. No.:	1,31e-43	Length:	834	
Score:	512.50	Matches:	116	
Percent Similarity:	63.57%	Conservative:	62	
Best Local Similarity:	41.43%	Mismatches:	91	
Query Match:	25.38%	Indels:	13	
DB:	29	Gaps:	2	

US-09-889-756a-2 (1-412) x CNS01JPK (1-834)

QY 106 SerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
 Db 831 GCAACCTTCAGCGCTTATGAAAGCCGAAAGCCGATCTGGCTTAAGACAGCCGCCG 772
 QY 126 LeuAlaLysAlaAspAlaAsp-LeuAlaArgTyrLysProLeuValAlaAlaGlnAlaVala 145
 Db 771 GHTAAATTTGCGCCMTGACGCTGAACCGCTATCAAAACGCTCGGTACGAAGTACAT 712
 QY 145 LSerArgGlnGlnTyrAspAlaAlaValAlaThrAlaLysArgSerAlaGlnAlaGlnVally 165
 Db 711 CAGTCAGAGGATTAACGATACCGCCCTGGCGGATGCCAGACAGTAAACGCGCGCTGCT 652
 QY 165 AAlaAlaGlnAlaAlaAlaLysSerAlaGlyLysAsnLeuAsnArgSerArgLysLeuAla 185
 Db 651 GCGACGCCAAAGCGCGCTGCAAAACCGCGCATTAACCTGCTATACAAAGTACCTC 592
 QY 185 AProLysSerGlyPheLysGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaG 205
 Db 591 CCTATACGCGCTGATTTGGTAAATCTTCCTCAGGAAAGGGGCTGTCGCAAGACGG 532
 QY 205 YAspThrThrValLeuAlaThrThrLeuArgGlnThrAsnProMetTyrValAsnValThrG 225
 Db 531 TCAGACCACTGCGCTGGCAACCGCGACACCTGATCCGATCTATGTGACGTCACGCA 472
 QY 225 nSerAlaSerGlnValMetLysLeuArgArgGlnLeuAlaGlnGlyLysLeuAlaAla 245
 Db 471 GTCCAGCAATATTCCTCGCGCTGAAACAGAGCTGCTAACGCGCCTGAAACAGGA 412
 QY 245 AAspGlyValAlaLeuAlaGlyLysPheAspAspGlyThrValTyrProGlnLysG 265
 Db 411 AAACGCAAA-GCCAAAGTGAGCGATTAACCAAGCGATTCAGAGTTCCCGCAGGAAG 353
 QY 265 YArgLeuLeuPheAlaAspProValValAsnGluSerThrGlnLysLeuAlaG 285
 Db 352 GACCTCGAATTCCTGACGTGACGTCGACCAACACCGCTTCATCACCCTTACGTGC 293
 QY 285 AAlaValProAsnAspGlnAsnLeuMetProGlyLeuTyrValArgValLeuMetAs 305
 Db 292 GATTTCCCGAACCCTGACAAAAATCTGCTGCCAGATGTTCTTGCGCAGCATCTGA 233
 QY 305 pGlnValAlaValAspAsnAlaPheValAlaProGlnGlnAlaValThrArg-----G 323
 Db 232 AGAAGGAACGAATCAACCGCCCTTCTGTTCCACAGAGGCTGACCCCTACCCACGC 173
 QY 323 YAlaLysAspThrValMetLeuValAsnAlaGlnGlyLysMetGluProArgGlnValTh 343
 Db 172 GCGCATGCG-AGCGACATGTTGTTGGGGCTGATTAACAAAGTCAAAATGCGCATATCAC 114
 QY 343 rValAlaGlnGlnGlnGlnThrAsnTyrPheValThrSerGlyLeuLysAspGlyAsp 363
 Db 113 CGCACCCAGCGGATTTGGGTAATGCTGTGACCGAAGTCTGAAGATGGCGATTCG 54
 QY 363 sValValValGlnGlyLysSerLeuAlaGlyLysThrGlyAlaLysLysValThrPro 382
 Db 53 CGTGATTTGTT-----ACTGGTTGCAAAAGTTCCTCT 20

RESULT 3
 BH368028 642 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-138C13.TR ND-TAM Anopheles gambiae genomic clone AG-ND-138C13
 DEFINITION
 ACCESSION
 VERSION BH368028.1 GI:17314072
 KEYWORDS GSS.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 REFERENCE 1 (bases 1 to 642)

AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carlisle, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.

TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*

JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)

MEDLINE 22542063

PUBMED 12655398

COMMENT Other GSSs: AG-ND-138C13.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@igr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source location/qualifiers
 1..642
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1, site_1: HindIII"

BASE COUNT 168 a 178 c 161 g 135 t

ORIGIN

Alignment Scores:
 Pred. No.: 3,97e-33 Length: 642
 Score: 411.50 Matches: 91
 Percent Similarity: 66.15% Conservative: 36
 Best Local Similarity: 47.40% Mismatches: 60
 Query Match: 20.38% Indels: 5
 DB: 28 Gaps: 1

US-09-889-756a-2 (1-412) x BH368028 (1-642)

QY 11 AAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyAsp 30
 Db 75 GCGCAGTTTGTAGTCTTTCAGCAGCTTAGCTTAACAGAGATTAACGATTAAGAAACC 134
 QY 31 AAlaAlaGlnGlyLysGlnProAlaGlyArgGlnAlaProAlaProValAlaGlyValAla 50
 Db 135 CAACAGAAAGCGCCCAAG-----GCACCGGCTCCCGAAGGGGTGATG 179
 QY 51 ThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuGluSerLeu 70
 Db 180 ACATTGAGAGCAACACTCTCAACATCACACCGATCTTCGTGCTGACCGCTGCTAT 239
 QY 71 ArgThrAlaAspValArgAlaGlnValGlyLysLeuValSerLeuValSerLeuVal 90
 Db 240 CGTATGCCGAAATTCCTCAGGTAGCGGATTAATCTGAACCTTAACCTTCGTAGAA 299
 QY 91 GlySerTyrValArgAlaGlnGlnProLeuTyrGlnLysSerSerThrTyrGlnAla 110
 Db 300 GCGACGAGATATCAAGCAGGTACTTCCCTATCAAAATGATCTGCTACTATCAGGCA 359
 QY 111 AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAsp 130
 Db 360 GCCTATGACAGCGCTTAAGCGGATTTGGCGAAAGCCCAAGCCAGCGCTATCGCAGCG 419

Oy 131 AlaaspleuAlaargTyrLeuProLeuValAlaAlaValSerArgGlnGluTyr 150
 Db 420 GTGACGGTTACCGCTTACCAACCATTTGCTGGTACCATGTTACATGATGACAAATTAC 479
 Oy 151 AspAlaAlaValAlaValSerArgSerArgValAlaValAlaAlaAla 170
 Db 480 GACACCGGAGTCTCCACCTGCAGCAGCGGAGCGGCTGTTGCTGCCAAAGCCGCT 539
 Oy 171 IleArgSerArgValIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
 Db 540 GTGAAACCGCGCGCATCAACCTGGCTTACACCAAGGTGACTTACCAATCTCCGCGCT 599
 Oy 191 IleGlyGlnSerValSerGlnGlyThrLeuLeu 202
 Db 600 ATCGCAATCTGCCCTGACCGAAGGCGACCTGTC 635
 RESULT 4
 LOCUS BH190105 601 bp DNA linear GSS 19-OCT-2001
 DEFINITION ATXOD17TF ATXO Arabidopsis thaliana genomic clone ATXOD17, genomic survey sequence.
 ACCESSION BH190105
 VERSION BH190105.1 GI:16302944
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (chale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 601)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Utechtack, T.V., Feldblyum, T.V. and Fraser, C.M.,
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 Unpublished
 Other GSSs: ATXOD17TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from some non-Arabidopsis source
 Seq primer: TP
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..601
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOD17"
 /clone_id="ATXO"
 /note="Vector: pPOS2; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pPOS2 using BstXI linkers"
 BASE COUNT 98 a 206 c 175 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.54e-29 Length: 601
 Score: 370.50 Matches: 72
 Percent Similarity: 66.11% Conservative: 47
 Best Local Similarity: 40.00% Mismatches: 58
 Query Match: 18.35% Indels: 3
 Gaps: 2
 US-09-889-756a-2 (1-412) x BH190105 (1-601)
 Oy 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 600 AAGGCGAGGTTCAGCAGAGGTGACTGTAGGCGCAGTATGAGACCAACCTGCTCCGCGTG 541

Oy 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232
 Db 540 ATCAACAATCCAGCCGATCTATTGACTTCAACGACGTGACCTCCGATCTCTGCGG 481
 Oy 233 LeuArgArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAsp--GlyValIleAlaVal 251
 Db 480 CTGAAGCGGGGAGTGAAGAGAGCGGAACCTTCCGAGGTGAGCCCGGAGACGCGCAGC 421
 Oy 252 GlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAlaAsp 271
 Db 420 AAGCTGCTGTGGAAGAGCGGACCGAGATTTCCAGACCGGCAAGATCTCTTCCGAA 361
 Oy 272 ProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGln 291
 Db 360 GCGGTGATGATGAGACGACCGGATGTACGCTGCGCGGAGTTTCCCAATCCGAG 301
 Oy 292 AsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsn 311
 Db 300 AAGATCCTGCTTCCCGGCATGTTGCGCGGTGCGGTGTCAGCGCGTGAAGAGAGAC 241
 Oy 312 AlaPheValValProGlnGlnAlaValThrArg-----GlyAlaLysAspThrValMet 329
 Db 240 GTGCTACCGGTGACGACGCGCGGTGACCGCATGCAAGCGGTGCGCGCAGCGTATG 181
 Oy 330 IleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGly 349
 Db 180 GTCATCAGCATTTCCACCAACGCGCAGATCCGATGATCCAGTCAACACGCGGTGGGC 121
 Oy 350 ThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValGluGlyIle 369
 Db 120 GACAAGTGGGTGTGACCTCGCGGCTCAAGGCGGAGAGAAAGCATCATGAGAGGCTTC 61
 RESULT 5
 LOCUS BZ558375 1096 bp DNA linear GSS 17-DEC-2002
 DEFINITION pacs1-60_968.s1 pacs1-60 Pseudomonas aeruginosa genomic clone pacs1-60_968, genomic survey sequence.
 ACCESSION BZ558375
 VERSION BZ558375.1 GI:27172643
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 1096)
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.,
 Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
 J. Bacteriol., (2002) in press
 Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.
 FEATURES
 source Location/Qualifiers
 1..1096
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pacs1-60_968"
 /clone_id="pacs1-60"
 /note="clinical isolate 1-60 Whole genomic shotgun library."
 BASE COUNT 181 a 385 c 351 g 179 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.35e-25 Length: 1096

Score:	343.00	Matches:	86
Percent Similarity:	55.84%	Conservative:	43
Best Local Similarity:	35.93%	Mismatches:	99
Query Match:	16.99%	Indels:	3
DB:	29	Gaps:	2

US-09-889-756A-2 (1-412) x BZ558375 (1-1096)

[illegible]

JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond

University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel.: 2062216954
Tel.: 2066857244
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES	Location/Qualifiers
source	1. .1503

```

/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8287"
/clone_1fb="pacs2-164"
/note="clinical isolate 2-164 whole genomic shotgun library"

```

[illegible]

```

Db 786 GTTAGCAGATCCGAC-TACTTGAACCTTTCAGCGNNGNG-----AAA 835
Qy 223 LEUARGAGGlnllealagluGlyys-----leuAlaAlaAspGlyValIle 249
Db 836 TTGGCGCATTCGAAMGGGATCCGGAAGAGGTGAAGTGTTCCACAAGAAATTN 895
Qy 250 AlaValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuPhe 269
Db 896 CCGTCGCGCTTCTCTTGGACNACCGAAGAAATACCCCTGTGNGCGGAGAGTGTTT 955
Qy 270 AlaAspProValValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsn 289
Db 956 TCGAACCTGGNGGTGAACCGGAGNACCGAACCATATGATGATGCGCGTTGNCGAAT 1015
Qy 290 AspinAsnIleLeuMetPro 296
Db 1016 ACCCATTTGAATTGTGCCC 1036

RESULT 7
B2550552/c 775 bp DNA linear GSS 17-DEC-2002
DEFINITION pasci1-60_2713.y2 pasci1-60 Pseudomonas aeruginosa genomic clone
ACCESSION B2550552
VERSION B2550552.1 GI:27154133
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 775)
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
AUTHORS Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-genome-Sequence Variation among multiple isolates of
JOURNAL Pseudomonas aeruginosa library
COMMENT J. Bacteriol., (2002) in press
Genome Center
Contact: Chris K. Raymond
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source Location/Qualifiers
1..775
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pasci1-60_2713"
/clone_lib="pasci1-60"
/notes="Clinical isolate: 1-60 whole genomic shotgun
library."
BASE COUNT 120 a 247 c 252 g 154 t 2 others
ORIGIN
Alignment Scores: 7.46e-23 Length: 775
Pred. No.: 315.00 Matches: 77
Percent Similarity: 53.33% Conservative: 43
Best Local Similarity: 34.22% Mismatches: 89
Query Match: 15.60% Indels: 16
DB: 29 Gaps: 6
US-09-889-756a-2 (1-412) x B2550552 (1-775)
Qy 192 GlnGlnSerLysValSerGlnGlyThrLeuAsnAlaGlyAspThrThrValLeuAla 211
Db 760 GGCGCTTCCTCCGCTTACGAGAGGCCCTGTGTACACCGCCGCAACCGGAATGCCC 701
Qy 212 ThrIleArgGlnThrAsnPro---MetTyrValaAsnValThrGlnSerAlaSerGluVal 230

```

```

Db 700 ACCGTGACACGAGCTTCGACCGATCTACGTCAACGTCACACCGCGTCCACCGCCCTG 641
Qy 231 MetLysLeuArgAGGlnllealagluGlyLysLeu---leuAlaAlaAspGlyValIle 249
Db 640 CTGGCGCTGNCCTCGAATGGGAGCGGCCAGTTGAGCGCGCGNCGACACCGCCGCG 581
Qy 250 AlaValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuPhe 269
Db 580 AAGTCTCCCTGAAGCTGAGAGACGTAGCCAAATACCCGCTGGAAGTCCGCTCGAATT 521
Qy 270 AlaAspProValValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsn 289
Db 520 TCCAGATTCCGTCGACGAGGACCGCGCTCGCTCAATCCGCGCTGTTCCCAAC 461
Qy 290 AspinAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaVal 309
Db 460 CCGAACACGAGCTGCTCGCGCATGTCTTACCGCGCAGTTGACAGAGGCGTCAAG 401
Qy 310 AspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----Thr 327
Db 400 CAGAGGCGCATCTCTCGCTCCGACGACAGCGCTGACCCGACCTCAAGCGCAGGTACC 341
Qy 328 ValMetIleValaAsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGln 347
Db 340 GCGCTGTGTGAACGCGGACGACACAGATGACGCTGCGGGGTGATCAAGCGCGAGTGT 281
Qy 348 GlnGlyThrAsnThrPheValThrSerGlyLeuLysAspGlyAspLysValValGlu 367
Db 280 ATCGGCGACAGATGCTGTGTTACCGAAGCGCTGAACCGCGGACGACAGATATTACCGAA 221
Qy 368 GlyIleSerIleAla-----GlyIleThrGlyAlaLysLysValThrProLysGluTyr 385
Db 220 GGCTCGCAGTTCGTCGACCGCGGTGTGAG---CTGAAGACCGTGGCGGCAAGAATGTC 164
Qy 386 AlaSerSerGluAsnGlnAlaAlaIleProGlnSerGlyValGlnThrAlaSerGluAla 405
Db 163 GCGTCCGCGGACGAGCGCGACGCGCTCG-----GCG 131
Qy 406 LysThrAlaSerGlu 410
Db 130 AAAACGACAGCAAG 116

RESULT 8
BHL90022 515 bp DNA linear GSS 19-OCT-2001
DEFINITION ATXOD32TF Arabidopsis thaliana genomic clone ATXOD32, genomic
survey sequence.
ACCESSION BHL90022
VERSION BHL90022.1 GI:16302757
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
1 (bases 1 to 515)
REFERENCE Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
AUTHORS Feldblyum,T.V. and Fraser,C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC T24G4
JOURNAL Unpublished
COMMENT Other GSSs: ATXOD32TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U cortig 1142. Caution: the DNA in this BAC may be from
some non-Arabidopsis source
Seg primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers

```

source 1. 515
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOD32"
 /note="Vector: pHOS2, Site 1: BstXI, 2-3 kb sheared BAC DNA inserted into pHOS2 using BstXI linkers"

BASE COUNT 97 a 164 c 173 g 81 t

ORIGIN

Alignment Scores:
 Pred. No.: 6.14e-22 Length: 515
 Score: 304.00 Matches: 71
 Percent Similarity: 64.24% Conservative: 35
 Best Local Similarity: 43.03% Mismatches: 53
 Query Match: 15.06% Indels: 6
 DB: 28 Gaps: 2

US-09-889-756a-2 (1-412) x BH190022 (1-515)

QY 20 leuValleuSerSerCysGlyLysGlyLysAlaAlaGlnGlyGlnProAlaGly 39
 |||:::|
 Db 38 CTGCTGCTGCCCGCTTGCCAAAGAGACAGCGGCC-----GGCCCGCG----- 82

QY 40 ArgGluAlaProAlaProAlaValAlaGlyValAlaThrValHisProGlnThrValAlaLeu 59
 ::::|
 Db 83 ---CAATGCGCGCTGCCCGCTGACCTCTCTGCGCGCGGACCGAAACGCTAGCATC 139

QY 60 ThrValGluLeuProGlyValGlyLeuGlnSerLeuArgThrAlaAspValAlaGlnVal 79
 |||:::|
 Db 140 ACCCGCAACTTCCCGCGCGGATCGATCGGTCGAGTGGGAGTGGCGCGCGCGCTC 199

QY 80 GlyGlyLeuIleGlnLysArgLeuPheGlnGlnGlnGlnSerTyValAlaAlaGlyGlnPro 99
 |||:::|
 Db 200 GCGCGCACTTCTGCGAGAGACCTTCCAGAGAGCGATCGATGAGCGGCGTGCAGCTG 259

QY 100 LeuTyGlnIleAspSerSerThrTyGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
 |||:::|
 Db 260 CTTTCAAGATCGATCCGCGCTCCGCTCGAGGCGCGGAGAGATGCGCGCGCTCTC 319

QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeu 139
 |||:::|
 Db 320 GCAGGCGCGGAGGAGACTTGCACAGAGCAAGCAGAGCTTATCGCTACGATCCCTG 379

QY 140 ValAlaAlaGluAlaValSerArgGlnGlnTyThrAlaAlaValThrAlaLysArgSer 159
 ::::|
 Db 380 GTGACTTCGAAACGGGTCAGCAAGCAGGCTTATGACATGCGGAGTCCGCGTGAAGTG 439

QY 160 AlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 |||:::|
 Db 440 GCGAGCGCCAGACTCAAGCGCGCGCGCGCGCTGAAGACCGCGCAACTGAACTGGGC 499

QY 180 ArgSerArgIleThr 184
 ::::|
 Db 500 TATCGACGGTGAACG 514

RESULT 9
 BZ554554/c 1236 bp DNA linear GSS 17-DEC-2002
 LOCUS pasci-60_4744.y2 pasci-60 Pseudomonas aeruginosa genomic clone
 DEFINITION pasci-60_4744, genomic survey sequence.
 ACCESSION BZ554554
 VERSION BZ554554.1 GI:27162165
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 REFERENCE 1 (bases 1 to 1236)
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of
 JOURNAL Pseudomonas aeruginosa library
 COMMENT J. Bacteriol., (2002) In press
 Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun

FEATURES
 Location/Qualifiers
 source 1. 1236
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pasci-60_4744"
 /note="pasci-60"
 /note="clinical isolate 1-60 Whole genomic shotgun library"

BASE COUNT 237 a 325 c 289 g 383 t 2 others

ORIGIN

Alignment Scores:
 Pred. No.: 4.21e-20 Length: 1236
 Score: 291.50 Matches: 77
 Percent Similarity: 50.93% Conservative: 32
 Best Local Similarity: 35.98% Mismatches: 97
 Query Match: 14.44% Indels: 8
 DB: 29 Gaps: 2

US-09-889-756a-2 (1-412) x BZ554554 (1-1236)

QY 13 AlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyLysAlaAla 32
 |||:::|
 Db 730 GCATTTTCAACCGATGATCGCTTGATTTGGTCAGCGGCGCACCTTAGAACGCTTC 671

QY 32 agnGlyGlyGlnPro-----AlaGlyArg-----GluAlaProAlaPr 45
 ::::|
 Db 670 CTCGAGCGGAGACCTTCGCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCT 611

QY 45 oValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThrValGluLeuProG 65
 |||:::|
 Db 610 CACNGTCGGGTGCGAAGGTGAGAGAGCGGACCTGCGCTCATTTCAACCGCTGG 551

QY 65 yArgLeuGlnSerLeuArgThrAlaAspValAlaGlnAlaGlnGlyIleGlnTy 85
 ::::|
 Db 550 CACCGTCACCGCTTCAACGCGTGAACGTCAGCGCGGCTCAACGCGAGCTGATCAA 491

QY 85 sArgLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 105
 |||:::|
 Db 490 GGTGCTGTTCCAGAGAGGCGGAGGAGGTCAGAGCGCGGACCTGCTGCGGTGTCACC 421

QY 105 rSerThrTyGlnAlaAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaTh 125
 |||:::|
 Db 430 GCGCACTTCAAGCGCGCGCTGCGCCAGCGCGGACGCTGATGCAGAACCGAGCGCA 371

QY 125 rLeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeuValAlaAlaGlnAlaVal 145
 |||:::|
 Db 370 ACTGAAGAAGCGCGAGATGACCTGCGCGCTCAAGGCGGCTGATGCGGAGACTGCAT 311

QY 145 lSerArgGlnGlnTyThrAlaAlaValAlaThrAlaLysArgSerAlaGluAlaGlyVal 165
 ::::|
 Db 310 CGCGAGCGAGACCTTGATACCCAGAGACCGGATCGCGCACTTGCAGGCGCACTCCG 291

QY 165 sAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAl 185
 ::::|
 Db 250 TACCAACAGCGCGGAGGTGACGAGACCGCGCTCACTGACCTTACCGAGTCCGCCG 191

QY 185 apPolleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaG 205
 |||:::|
 Db 190 ACCGATTTCCGGCGCGCTGCGCTACGCGAGGTGACATCGGCAACTGTGTCACGCGG 131

Oy 205 YAspThrValIeuAlaTrnIleArgGlnThrAsnPro 218
 Db 130 CGATACCAAGCGCGCTGTGTGATACACCGAGTCAAGCG 91
 RESULT 10
 BH250378 798 bp DNA linear GSS 26-NOV-2001
 LOCUS BOCAB19TF BOGA Brassica oleracea genomic clone BOCAB19, genomic
 DEFINITION survey sequence.
 ACCESSION BH250378
 VERSION BH250378.1 GI:17074104
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 798)
 Town, C.D., Van Aken, S., Uteerback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 JOURNAL Other_GSSs: BOCAB19TR
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..798
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGAB19"
 /note="Vector: pHOSt1, Site 1: BstXI, 2-3 kb sheared
 genomic DNA inserted into pHOSt1 using BstXI linkers"
 BASE COUNT 149 a 246 c 238 g 165 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,766-19 Length: 798
 Score: 280.00 Matches: 66
 Percent Similarity: 62.88% Conservative: 17
 Best Local Similarity: 50.00% Mismatches: 41
 Query Match: 13.87% Indels: 8
 DB: 28 Gaps: 1
 US-09-889-756a-2 (1-412) x BH250378 (1-798)
 Oy 16 AlaAlaValAlaLeuValIeuSerSerCysGlyGlyGlyAlaAspAlaGlnGlyGly 35
 Db 425 GCGACGGCTGCCGTTTCTCTGCGCGCGTGGGAAA----- 460
 Oy 36 GlnProAlaGlyArgGlnAlaProAlaProAlaValGlyValValThrValHisProGln 55
 Db 461 AAAGATCGCGCTCCGCCCTCAAGCGCCGAAAGTCGGGTGTCACCGTCACGCGAG 520
 Oy 56 ThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspVal 75
 Db 521 CCGGCGCGGTCTCGCAACTGCGGCGCGCTACGAGCGCTATCTGCGCGAGGTG 580
 Oy 76 ArgAlaGlnValGlyGlyIleGlnIleArgLeuPheGlnGlnGlySerThrValArg 95
 Db 581 CCGCGACGGGTGACGATCTGTGGCGCGCGCTTACGGAAGGACGACGCTCAAG 640
 Oy 96 AlaGlyGlnProLeuThrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAla 115

Db 641 GCCGCGACGGCGCTGTACAAATCGATCCGGACCCCTATATCGCGCAATTGAACAGCGCG 700
 Oy 116 ArgAlaGlnLeuAlaTrnIleArgGlnAlaThrLeuAlaIleAspAlaAspLeuAlaArg 135
 Db 701 AAGCGACGCTCGGGAAGCGCGGACCACTCCGACGACGAGAACGCGCTGTGCGCGCG 760
 Oy 136 TyrIleProLeuValAlaAlaGlnAlaValSerArg 147
 Db 761 TACAAAGTGTCTGTCGCGCGCAACGCGGTGACGAG 796
 RESULT 11
 BH189811/c 620 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOC54TF ATXO Arabidopsis thaliana genomic clone ATXOC54, genomic
 DEFINITION survey sequence.
 ACCESSION BH189811
 VERSION BH189811.1 GI:16302308
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 620)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uteerback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 Survey sequencing of Arabidopsis thaliana BAC T24G4
 Unpublished
 JOURNAL Other_GSSs: ATXOC54TR
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from some non-Arabidopsis source
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..620
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATXOC54"
 /note="Vector: pHOSt2, Site 1: BstXI, 2-3 kb sheared BAC
 DNA inserted into pHOSt2 using BstXI linkers"
 BASE COUNT 108 a 191 c 199 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,46-18 Length: 620
 Score: 269.50 Matches: 65
 Percent Similarity: 61.29% Conservative: 30
 Best Local Similarity: 41.94% Mismatches: 47
 Query Match: 13.35% Indels: 13
 DB: 28 Gaps: 3
 US-09-889-756a-2 (1-412) x BH189811 (1-620)
 Oy 20 LeuValIeuSerSerCysGlyGlyGlyAlaAspAlaGlnGlyGlnProAlaGly 39
 Db 430 CTGCTGCTGCCGCTTGCAAAAGACAGACGCGC-----GGCCCGCC----- 386
 Oy 40 ArgGlnAlaProAlaProAlaValGlyValValThrValHisProGlnThrValAlaLeu 59
 Db 385 ---CAATCGCGCTGCGCGCGGTGACCTTCTCGCGCGGTGACCGAAGCGTACGATC 329
 Oy 60 ThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValAlaGlnVal 79
 Db 328 ACCGCGAACTTCCGCGCGGATGATCGGTGAGTGGCGAGGTGCGCGCGCTC 269

Qy 80 GlyGlyIleIleGlnIleuArgLeuPheGlnGlySerTyrValArgAlaGlyGlnPro 99
Db 268 GCGCGCATCTCTGCTGAGAAAGACTTCCAGGAAGGCATGACGTAAAGCGCGGTACCTG 209
Qy 100 LeuTyrGlnIleAspSerSerThyTyrGlnIleAsnLeuGluSerAlaArgAlaGlnLeu 119
Db 208 CTTTCAAGATCGATCCGCTCCGCTCCAGCGCGGAGAAATGCGCGCGCTCTC 149
Qy 120 AlaThrAlaGlnAlaThrLeuAlaIleValAlaAspLeuAlaArgTyrLeuProLeu 139
Db 148 GCAAGGCGCGAGGCCAAGCTTGCACGAGCAAGACGACGCTGATCGCTACGAGTCCCTG 89
Qy 140 ValAlaAlaGlnAlaValSerArgGlnIleuTyrAspAlaAlaValThrAlaLeuArgSer 159
Db 88 GTGACTTCGAACGCGGTCCAGCAGCAGCGCTTATGAC-----AAT 50
Qy 160 AlaGlnAlaGlyValIleValAlaAlaGlnAlaAlaIleuSerAla 174
Db 49 GCGAGTCCGCGGTAAAGTGGCCGACCGACGCTCAAAAGCGCG 5
RESULT 12
LOCUS BZ557583 1263 bp DNA linear GSS 17-DEC-2002
DEFINITION pacel1-60_635.s2 pacel1-60 Pseudomonas aeruginosa genomic clone
pacel1-60_635, genomic survey sequence.
ACCESSION BZ557583
VERSION BZ557583.1 GI:27170544
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1263)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence Variation among multiple Isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
FEATURES
source Location/Qualifiers
1..1263
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacel1-60_635"
/clone_1db="pacel1-60"
/note="clinical isolate 1-60 whole genomic shotgun
library."
BASE COUNT 208 a 386 c 412 g 257 t
ORIGIN
Alignment Scores:
Pred. No.: 6.26e-17 Length: 1263
Score: 261.50 Matches: 96
Percent Similarity: 41.72% Conservative: 45
Best Local Similarity: 28.40% Mismatches: 125
Query Match: 12.95% Indels: 72
DB: 29 Gaps: 9
US-09-889-756a-2 (1-412) x BZ557583 (1-1263)
Qy 69 SerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnIleuArgLeuPhe 88
Db 37 TCGTTCACAACGTGTGATCCGACCCAGATGACGCGCATGACCCGCTGCTG 96

Qy 89 GlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThyTyr 108
Db 97 AGCGAAGGCGAGATGCTCGAGCGCGGCGAGTGTCTGGCGACCATCGATGACCGCGCTC 156
Qy 109 GlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 157 GTCCCGCGCTGAGCAGCGCGCGAGGCTCTGCGCGAGCTACCGAGCCAGCTGATATCC 216
Qy 129 AlaAspAlaAspLeuAlaArgTyrLeuProLeuValAlaAlaGlnAlaValSerAlaGln 148
Db 217 GCCAGAGAACCTCGCAACCTTACCGCAGCTGTATCCAGCGCTGCTCTATGCCAA 276
Qy 149 GluTyrAspAlaAlaValThrAlaLeuArgSerAlaGlnAlaGlyValIleAlaAlaGln 168
Db 277 CTGTGTGACTACGACGAGCGCGCGCTGACCATGTTGGCGCTCTGAAAGCCAAAGAT 336
Qy 169 AlaAlaIleuSerAlaGlyIleAsnLeuAsnArgSerAlaThrAlaProIleSer 188
Db 337 GCCACCATCAACCGCGAGCGGCTGCGCTGTCTTACACCGGATCATCTCGCGGTATCC 396
Qy 189 GlyPheIleGlyGlnSerArgValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThr 208
Db 397 GGCAGGTCTGTATCCGCAACGTGATGTGGCAACCTGTGCGGTCGCGACAGTCTC 456
Qy 209 ValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSer 228
Db 457 GGCCTGTTCAGCGTACCGACCATGATGACCATCTCTCTTCTCTCCGCAACAGAA 516
Qy 229 GluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyVal 248
Db 517 CAGTGGCCCGAGTTGCAG-----GCCGTGCTCGCGCGGCGACGCGGCG 558
Qy 249 Ile-----AlaValGlyIleLysPheAspGlyThr 259
Db 559 GTGCGCGCTTACGACCGCGACCGCGGAGCGCGCTGGCC-----597
Qy 260 ValTyrProGlnLysArgLeuLeuPheAlaAspProValAlaAsnGlnSerThrGly 279
Db 598 -----CGAGGCGGTTGCGACCATGCAACCAAGATCGCAATTTCCACGCT 645
Qy 280 GlnIleThrLeuArgAla-----AlaValProAsnAspGlnAsnIleLeu 294
Db 646 ACCATCGCGGTGGCGCGCTCTCTGATGATACCGCGCAGCGCGC-----TTC 693
Qy 295 MetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAlaPheVal 314
Db 694 TTGCCCCGCGAGTTCGCTCGCGCGCTTACGTTCA-----729
Qy 315 ValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGln 334
Db 730 -----CACCGGCGCTC-----741
Qy 335 GlyGlyMetGlnProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTyrIleVal 354
Db 742 ---GGCGGGAACCA-----GTTGGTCTGTGCGGACCAAAAGCGTTGCG 786
Qy 355 ThrSerGlyLeuLysAspGlyAspLysValValGlnGlyIleSerIleAlaGlyIle 374
Db 787 GCGCGCGCGCTTGTGGAAGGAATTCCTGCTTACGCGGGGTTACGCTCCCGACCA 846
Qy 374 eThrGlyAla-----LysLysValThrProLysGln 384
Db 847 CACGGGTCTCTAAAGCGCGGTTGGCCAAAGTTCCACGCGTTGCTCCCAAGGAG 898
RESULT 13
LOCUS BZ562652 937 bp DNA linear GSS 17-DEC-2002
DEFINITION pacel2-164_3863.x1 pacel2-164 Pseudomonas aeruginosa genomic clone
pacel2-164_3863, genomic survey sequence.
ACCESSION BZ562652
VERSION BZ562652.1 GI:27185018
KEYWORDS GSS.

SOURCE
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 3937)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press

JOURNAL
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..937
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_3863"
/clone_lib="pacs2-164"
/note="Clinical isolate 2-164 Whole genomic shotgun
library"

BASE COUNT
155 a 303 c 327 g 147 t 3 others

ORIGIN

Alignment Scores:
Pred. No.: 2,566-16 Length: 937
Score: 254.00 Matches: 91
Percent Similarity: 42.09% Conservative: 42
Best Local Similarity: 28.80% Mismatches: 132
Query Match: 12.58% Indels: 51
DB: 29 Gaps: 8

US-09-889-756a-2 (1-412) x BZ562652 (1-937)

Qy 99 ProLeuTYrGlnIleAspSerSerThrTYrGluAlaAsnLeuGluSerAlaArgAlaGln 118
Db 93 CCTGTTACCCCATCGATGATGACCGCGCGCTGCGCGCTGAGCAGGCTCAGCGCTCC 152
Qy 119 LeuAlaThrAlaGlnAlaThrLeuAlaAlaAspAlaAspLeuAlaArgTYrIleSpro 138
Db 153 AGGCGGACCAACCAAGCCAGCTGAATCCGCGGACGAGACCTGCAACGCTACCGCAGC 212
Qy 139 LeuValAlaAlaGluAlaValaSerArgGlnGluTYrAspAlaAlaValaThrAlaIleSArg 158
Db 213 CTGTATGCGCAGCGCGCGGTGTCGGGCCAAGCTGCGACGACGACGACGACGCGTCGAC 272
Qy 159 SerAlaGluAlaGlyValaIleAlaAlaAlaIleSeraIleGlyTYrLeuSleu 178
Db 273 CAGTTGCCGCGACCGCTGAAGGCCCAACATCAACGCGCAGCGCGGCTGCGCTG 332
Qy 179 AsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerIleValaSerGlu 198
Db 333 TCTTACACCGGATCATCTCGCGGTATCCGCAAGTCCGATCCGCAACGTCGATGTC 392
Qy 199 G1YThrIleLeuAsnAlaG1YAspThrThrValaIleuAlaThrIleArgG1nThrAsnPro 218
Db 393 GGCACACTGCGTGGCGGTGCGGACAGTCTCGGCTCTTACAGGTACACCAATCGCACCG 452
Qy 219 MetTYrValaAsnValaThrGlnSerAlaSerGluValaMetIleSeraIleArgGlnIle--- 237
Db 453 ATCTCCAGTACTCTCTCCCTGCAACAGACAGAGTGTCCCGCAGGTCGCGCTGCGC 512
Qy 238 -----AlaGluGlyIleLeuLeuAlaAlaAspGlyValaIleAlaValaG1YIleSphe 255
Db 513 GCGGAGGCGGCGGTGCGGCTCAACGCGCAGCGCGGAGGAGCGCGCTGCGC----- 563

Qy 256 AspAspGlyThrValTYrProGluIleG1YArgIleuLeuPheAlaAspProValaAsn 275
Db 564 -----GAGGCGCGGTGCTGCATCATGACCAACACAGATCGAC 599
Qy 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValaProAsnAspGlnAsnIleLeuMet 295
Db 600 AGTTCCACCGGTATCATCTCGCGGTGCGGCTCTTCATTAACCCGCGCGGCTCTGCG 659
Qy 296 ProGlyLeuTYrValaArgValaIleuMetAspGlnValaAlaAspAlaAlaPheVala 315
Db 660 CCGACGAGTGTNNCGG----- 677
Qy 316 ProGlnAlaValaThrArgGlyAlaIleAspThrValaMetIleValaAsnAlaGlnGly 335
Db 678 -----TGAGCTCAACCGCGGCGGTGCGGCGCACAGTGTGCTGTGACCAAGCGGG 731
Qy 336 GlyMetGluProArgGluValaThrValaAlaGlnGlnGlnIleTYr-----AsnTr 352
Db 732 GCGCGCGCGCTGAGGCAATTCGTTTACCGGCTGCGCGCGCGGTCGCGGCGGCTG 791
Qy 352 P-----IleValaThrSer-----GlyLeuIleAspGlyAspIleSVala 364
Db 792 GCCAGCGCGGTCTTACGACATGACGCGCTAAGTGTGTCGAAGCCCGCGCAGCGG-- 849
Qy 364 IValaIleGluGlyIleSerIleAlaGlyIleThrGlyAlaIleSValaValaThrProIleSgl 384
Db 850 -----GGACCGGCGGCGGTGACGCGCCATCCCGGTGATGATCCCGGCGG 893
Qy 384 UTPrAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyVala 399
Db 894 CTGTGTC-----AATATCAAGACCCCGCCCAACTGTGTTTA 930

RESULT 14
BZ574393 1417 bp DNA linear GSS 17-DEC-2002
LOCUS msh2_3648.x1 msh Pseudomonas aeruginosa genomic clone msh2_3648,
DEFINITION
BZ574393
ACCESSION BZ574393.1 GI:27209454
VERSION
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 1417)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press

JOURNAL
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..1417
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSh"
/db_xref="taxon:287"
/clone="msh2_3648"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library"

BASE COUNT
242 a 399 c 376 g 397 t 3 others

ORIGIN

Alignment Scores:
Pred. No.: 1,196-15 Length: 1417

Score:	250.00	Matches:	79
Percent Similarity:	46.83%	Conservative:	39%
Best Local Similarity:	31.35%	Mismatches:	11
Query Match:	12.38%	Indels:	18
DB:	29	Gaps:	4

US-09-889-756A-2 (1-412) X BZ574393 (1-1417)

QY 39 GYArgGlnAlaIleProValAlaGly-----ValAlaThrValHis 53
Db 37 GGTACGGGGCCCCCTCGAGGCTCCAGGATACGATTAAGCTTATATGAAATCTCTCAG 96
QY 54 ProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAla 73
Db 97 CCCCACAGGTCCAGCGGCATC-----GGCAGGTGATCCTGTTTCCAGCGCTG 144
QY 74 AspValArgAlaGlnValGlyGlyIleLeuGlnValArgLeuPheGlnGlnGlySerTyr 93
Db 145 GTGATCGCACCCACAGTCCAGGTCATGTCAGCCGCGCTGCGTAGAGGAGGACATG 204
QY 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGln 113
Db 205 GTGAGGCGGGGCGAATTGCTGCGGACCATTCATGACCGCGCGCTGCTGCGCTGGAG 264
QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeu 133
Db 265 CAGGCGGAGGCTTCGAGGGCGAGTACACAGGCGCAGCTGAATCCGCGGACAGACCTG 324
QY 134 AlaArgTyrIysProLeuValAlaAlaGlnAlaValSerArgGlnGluTyrAspAlaAla 153
Db 325 CAACGGTACCGCGCTGTATGCTCCGAGCGTGGGTGTACACCCAACTGCTGGACCAAG 384
QY 154 ValThrAlaIysArgSerAlaGlnAlaGlyValIleValAlaAlaGlnAlaIleIysSer 173
Db 385 CAGGCGACGGTGCACCACTGTCGGCGCGACCTCGAAAGGCCAAACAGTCCACCATACGCC 444
QY 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln 193
Db 445 GACGGGCTGGCTGTACTACACCCGGATCACTCGCGGTATCCGGCAGAGTGGGTATC 504
QY 194 SerIysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213
Db 505 CGCAACGTGATGTCGCAACTGTGGTGGCTGCGCGACATCTCGGCTGTTCACGGTG 564
QY 214 ArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGlnValMetIysLeu 233
Db 565 ACCCAATGCGACCATCTTCTGTAGCTTATCTTGCACAGCAAGACATTGCCCACTTG 624
QY 234 ArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValaGlyIle 253
Db 625 CAGGCGCTGCTGGC---GGAAAGCGGGGCTGCGGCTTACAGCGGTAACGNGGAGCG 681
QY 254 LysPheAspAspGlyThrValTyrProGluIysGlyAlaGlyLeuLeuPheAlaAspProVal 273
Db 682 CGCTT-GACGAG-----GCCCGGTGTCCACCATTTGCAACACAG 719
QY 274 ValAsnGluSerThrGlyGlnIleThrLeuArgAla 285
Db 720 ATGACAGATTCCACCGGCACTTCGCGGTGGGCGCC 755

RESULT 15	BF663822/c	802 bp	mRNA	linear	EST 21-DEC-2000
LOCUS	602145261F1				
DEFINITION	602145261F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4308951 5', mRNA sequence.				
ACCESSION	BF663822				
VERSION	BF663822.1	GI:11937717			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 802)				

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

found through the I.M.A.G.E. Consortium/LLNL at: <http://www.llnl.gov/ims>

FEATURES

Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4308951"
/tissue_type="primary B-cells from consils (cell line)"
/lab_host="DH10B (phage-resistant)"
/clone_idb="NIH_MGC_48"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAC(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

```

Alignment Scores:

Pred. No.:	6.97e-16	Length:	802
Score:	249.00	Matches:	87
Percent Similarity:	54.84	Conservative:	49
Best Local Similarity:	35.08*	Mismatches:	75
Query Match:	12.33*	Indels:	38
DB:	10	Gaps:	6

US-09-889-756A-2 (1-412) X BF663822 (1-802)

QY	30	AspAlaalaGlnGlyGlnProIlaGlyArgGlnIaProValaGlyVal	49
Db	726	GATGTGGCCGACCTGACAGTCCCTCTTCTAGTGGCCATCAGATGCCCTCTGTGGCCGTA	667
QY	50	ValThr---ValHisProGlnThrValAlaLeuThrValGlnIleuProGlyArgLeuGln	68
Db	666	GTGACCAAGTCTGT-CTGAACCTCTCCAGTACCAACGAGCTTCCGGGTGGACCAAG	608
QY	69	SerLeu-ArgThrAlaSerValArg-AlaGlnValGlyGlyIleIleGlnIleArgLeuP	88
Db	607	GCGTTACCGATCGCGAAGTTCGTCCTCAAGGTACGGGATTCATGAAAGGTAAGT	548
QY	88	heGlnGlnIySerYrValArgAlaGlyGlnProLeuYrGlnIleAspSerSerThrT	108
Db	547	TTTCAAGAAAGGTATCGACATCGAAGCAAGTGTCTCTTATCAGATTGATCTTGGACCT	488
QY	108	YrGlnAlaSerIleuGlnSerAla-	120
Db	487	ATCAGCGCATACAGCAGTGTGGAATAGTGATCTGGCGAAAGTCCAGGCTCAGGCTA	428
QY	120	IaThrAlaGlnIaThrLeuAlaLysAlaAspAlaSerLeuAlaArgYrGlySerProLeuY	140
Db	427	TATCGCGCATGTGTGCG-----GTGAATCGTTATCAGCAACTGC	389
QY	140	AlaAlaAlaGlnIaValSerArgGlnGlnYrAsp---AlaAlaValThrAlaLysArgS	159
Db	388	TGCGTACTTACATCACTAGTAAAGCAAGATGTCAGTACAGTCTCTGGCTGATGTCCGAAT	329

QY 159 eralaglu-Alaglyvallyalalaglin-Alalalaleyseralaglylleasme 178
 Db 328 CAGCGGATGCTGGCGATCTGCGCGAATGCTCGCTGAATCGCGCATCAATCT 269
 QY 178 uasmarserarg-ilethrallaprolleserglyphenileglylmerlyvalaser- 197
 Db 268 GCGCTTACACCAATGACCTCTCCGATTACCGGATGGTGAAGTGAACGAGAG 209
 QY 198 --gluglyThrleuenuasnalaglyasethrThrValleuAlathrillearglinThra 217
 Db 208 GAAGCGCATCTTGACAGTACGATGCGATGCGCTGCGCAACCTGACAGCACTTG 149
 QY 217 snpmetcylvalasenvalthrinserralaserglivalmetlysleuarg-gin 236
 Db 148 ATCCGATCTACGTGATGATGACCCAGTCCAGCAACGATCTCTCGGCTGAACAATGAA 89
 QY 237 llealaglugly-----lysleuenu 243
 Db 88 CTGCGCAATGGCAGCTGAATAATCAAGATGACGCGCAATAGCCACAGTGTATCTGATC 29
 QY 244 Alalalaspgly 247
 Db 28 ACCAGTGCAGCG 17
 RESULT 16
 BH190195 620 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXO492TR ATXO Arabidopsis thaliana genomic clone ATXO492, genomic
 DEFINITION survey sequence.
 ACCESSION BH190195
 VERSION BH190195.1 GI:16303143
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (chale crees)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 620)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Utecherback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 TITLE Survey sequencing of Arabidopsis thaliana BAC T24G4
 JOURNAL Unpublished
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seg primer: TR
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..620
 /organism="Arabidopsis thaliana"
 /mol type="genomic DNA"
 /strain="Columbia"
 /db xref="taxon:3702"
 /clone="ATXO492"
 /clone_id="ATXO"
 /note="Vector: pHO52; site 1: BstXI; 2-3 kb sheared BAC
 DNA inserted into pHO52 using BstXI linkers"
 BASE COUNT 113 a 201 c 184 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.01e-15 Length: 620
 Score: 246.00 Matches: 52
 Percent Similarity: 61.07% Conservative: 28
 Best Local Similarity: 39.69% Mismatches: 47
 Query Match: 12.18% Indels: 4
 DB: 28 Gaps: 3

US-09-889-756a-2 (1-412) x BH190195 (1-620)
 QY 265 G1Yargleuenupealaaasprovalvalasnuiserthrnglylinlethrleuarg 284
 Db 4 GCGAAGATCTCTCTTCCGAAAGCGGTGTGATAGACGACCGGATGATCAACCTGGC 63
 QY 285 Alalalavalproanaaspglnasnilleuuecproglyleuylvalargvalleuuec 304
 Db 64 GCGAGATTCCTCCATCCCGAAGATCTCTCTCCGCGCATGTTCGCGCGCGCGGTG 123
 QY 305 Aspglnvalalavalaspaalalphevalvalproglnglnalvalthrarg----- 322
 Db 124 GTGAGAGCGGTGAAGGAGCGGTGTCACCGTGCACAGCGCGCGTGAACCGGATCGAA 183
 QY 323 G1Yalalaysasprthrvalmetilevalasnalaglnlylgmetgluproargluval 342
 Db 184 GCGGTGCGCGGACGCGATGTGATGACAGATTCCAACGCGGACAGATCGGATGATC 243
 QY 343 ThrvalalaglnnglnlythrAsnTrpIlevalthrserglyleuylaspglyasp 362
 Db 244 CAGACTGACAAACGCGGTGGGACAGAGTGGGTGACCTCGGCTCAAGCGCGAGAG 303
 QY 363 Lysvalvalvalaglylylleserillealaglylethrghyalalysvalthrpro 382
 Db 304 AAAGTCATCATCGAAGGCTC--CTGAAGCTCTGTCGCGGAGCG--CCGCTGCTTCG 357
 QY 383 Lysglutrpilaserserghuasnglnalala 393
 Db 358 GAGCCCTTCGTCGCGAAGCGGACCGCTCG 390
 RESULT 17
 BH189826 596 bp DNA linear GSS 19-OCT-2001
 LOCUS ATXOC46TR ATXO Arabidopsis thaliana genomic clone ATXOC46, genomic
 DEFINITION survey sequence.
 ACCESSION BH189826
 VERSION BH189826.1 GI:16302338
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (chale crees)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 596)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Utecherback, T.V.,
 Feldblyum, T.V. and Fraser, C.M.
 TITLE Survey sequencing of Arabidopsis thaliana BAC T24G4
 JOURNAL Unpublished
 COMMENT Other GSSs: ATXOC46TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1142. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seg primer: TR
 Class: sheared ends.
 FEATURES
 source Location/Qualifiers
 1..596
 /organism="Arabidopsis thaliana"
 /mol type="genomic DNA"
 /strain="Columbia"
 /db xref="taxon:3702"
 /clone="ATXOC46"
 /clone_id="ATXO"
 /note="Vector: pHO52; site 1: BstXI; 2-3 kb sheared BAC
 DNA inserted into pHO52 using BstXI linkers"
 BASE COUNT 104 a 183 c 195 g 114 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.01e-15 Length: 596
 Score: 246.00 Matches: 52
 Percent Similarity: 61.07% Conservative: 28
 Best Local Similarity: 39.69% Mismatches: 47
 Query Match: 12.18% Indels: 4
 DB: 28 Gaps: 3

Pred. No.:	1.56e-15	length:	69
Score:	244.00	Matches:	51
Percent Similarity:	58.94%	Conservative:	28
Best Local Similarity:	40.40%	Mismatches:	48
Query Match:	12.09%	Indels:	14
DB:	28	Gaps:	3

US-09-889-756A-2 (1-412) x BH189826 (1-596)

```

QY 20 LeuValIenSerCysGlyIysGlyIysGlyAlaPheAlaAlaGlnGlyGlyGlnProAlaGly 39
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 424 CTCTGCTGCGCCCTTGCCAAAGAGCAAGCGCC-----GCGCCGCG-----380

QY 40 ArgGlnAlaProAlaProValAlaGlyValAlaThrAlaHisProGlnThrValAlaLeu 59
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 379 ---CAAATGCGCGCTGCGCGGATGACTCTTCCTGCGCGCGAGTACCGAAAGAGATGATGATC 323

```

Db 322 ACCCGGAACTTCCCGCGGATCGATCCGGGCGAGTGGCGGAGGAGCGTGCCTC 263

D_b 262 GGCGCATCTGCTGGAGAACCTTCCAGGAAGCATGTGAAGCGGGTGACGCTG 2033

Db 202 CTTTCAGATCGATCCGGCTCCGCTCGAGCGCGGAGAGATGCGCGGCGCTCTC 143

Db
142 GCACGGGCGGAGCACTTGACACGAGCAAGACGACCTTGATCGCTACGAGTCCCTG 83

```

Db      82 GTGACTTCGAACGGCGGTACAGCAAGAGGCTTAT-----GCA 47
      |||  ::: |||||:::|||||

```

Db 46 ATCGAAGAACGGTGAAGTGGCCGACGCGAGC 14

LOCUS	1108 bp	DNA	linear	GSS 17-DEC-200
DEFINITION	B2577818			
	msh2_5586.y2	msh	<i>Pseudomonas aeruginosa</i>	genomic clone msh2_5586,
	genomic survey sequence			

VERSION	B2577818.1	GI:27212879
KEYWORDS	GSS.	
SOURCE	Pseudomonas aeruginosa	

REFERENCE
1 (bases 1 to 1108)
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

TITLE	Author	Journal	Year
Burns, J.L., Kaul, R. and Olsen, M.V. Whole-Genome-Sequence variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library			

COMMENT:
Contact: Chris A. Raymond
Genome Center
University of Washington
Box 353145 Seattle WA 98105-3145 non

Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.

```

FEATURES
source
1.1108
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"

```

	/clone="msb2_5586"	
	/clone_lib="msb"	
	/note="Environmental isolate. Whole genomic shotgun library"	
BASE COUNT	242 a	352 c 306 g 192 t 16 others
ORIGIN	.	

Alignment Scores

Score:	241.50	Matches:	80
Percent Similarity:	56.63%	Conservative:	31
Best Local Similarity:	40.82%	Mismatches:	79
Query Match:	11.96%	Indels:	11
DB:	29	Gaps:	3

Db
Oy 19 LeuLlaIaAlaValLeuValLeuSerCysGlyLysGlyAspAlaAlaGln 33
||||| ||||| |||
230 CTGGCGGCCCTGTTCGCCCAATTCTCTGCTGGGCTGC-----GAAGAAGACGGGAC 340

Db 341 ACCGGGAAACTGCG-----GAGGCCCGCCCGAG--GTCCGGCGATGTCGCCAGG 391

Db 392 CCGGCGCTATTCGGCATCACAGCGAGCTCCCGGACGCTGGAGCGTACCGCCAGGCT 451

Db 452 GAAGTGGGGCGCGCGTCCGCGGATGTCAACCCGTGCTGTACGAGAGGCCAGGAC-510

D5 511 GTCCGCGCGGACCGCTGTTCCAGATGACCCCTGGCC-CTTGAAGCGGNCCTGAC 569

Db 570 ATCAGCGCGCGCCCTGGCCCGGCGCCGAGGCCAGNCA CGNNCGCGCGGNCACAAGCTC 629

Db 630 AGCCGCTACGCCGACCTGATCAAGSAGCGNGCCATCAGCGAAGCGGAGTACACCGAAGCG 689

Db 690 CAACCGACGGCGCAGCCCTGNNGCANATTGCTTGNNCAGAGCGGAA -CTTGACCAA 748

Db 749 GCCCGCCTGCG-CTTGGCTACGCCACGGTAAAGGGCCCCCATTTGACGACGACGCGGGTNT- 806

Oy 194 SerLysValSerGluGlyThrLeuLeuAsnIagiAspThrThrVal 209

D5 807 TCGCTGTACCAAGGCCCCCTTG6CCGGAAAGATTCCCGAACCGTT 854

LOCUS	BZ580698	1259 bp	DNA	linear	GSS 17-DEC-2000
DEFINITION	msh2_936.y2 msh Pseudomonas aeruginosa genomic clone msh2_936, genomic survey sequence.				
ACCESSION	DZ590006				

ORGANISM	<i>Pseudomonas aeruginosa</i>
SOURCE	<i>Pseudomonas aeruginosa</i>
KEYWORDS	GSS.
VERSION	B230030.1 G1.2/2135

REFERENCE
Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1259)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE
Whole-genome-Sequence variation among multiple isolates of

Db	229	CMAACCCGTGTTAAATTAACGAAACCCCTTACCGGAGGCTTTAAACAATGCGAAAGG	288
Qy	118	GInLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaSerLeuAlaArgTyrLys	137
Db	289	AGTTACATGCGACGGAAACCCCGTAAACCGCTACGTTAGAGGTTATAGTAACA	348
Qy	138	ProLeuValAlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLys	157
Db	349	CCACTGTTCACAAATTAAGTTGTTGTGATTCAGTTAAACCGCAAAACAGCTTAT	408
Qy	158	ArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsn	177
Db	409	AAATTTGGCAGCGCAAAATCGCAACAAGCTTAAGCAGTGTCTGCACACAGATTAC	468
RESULT 21			
LOCUS	BZ554963	923 bp	DNA linear GSS 17-DEC-2001
DEFINITION	pacal-60_4947.x1 pacal-60 Pseudomonas aeruginosa genomic clone		
ACCESSION	BZ554963		
VERSION	BZ554963.1	GI:27163258	
KEYWORDS	GSS.		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
AUTHORS	1 (bases 1 to 923) Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.		
TITLE	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library		
JOURNAL	J. Bacteriol., (2002) In press		
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 20622316954 Fax: 2066857244 Email: ckraymond@u.washington.edu Class: shotgun.		
FEATURES	Location/Qualifiers		
source	1..923 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="1-60" /db_xref="taxon:287" /clone="pacal-60_4947" /clone_1ib="pacal-60" /note="Clinical isolate 1-60 whole genomic shotgun library."		
BASE COUNT	190 a	256 g	212 t
ORIGIN	1 others		
Alignment Scores:			
Pred. No.:	1.08e-12	Length:	923
Score:	219..50	Matches:	66
Percent Similarity:	50.98%	Conservative:	38
Best Local Similarity:	32.35%	Mismatches:	89
Query Match:	10.87%	Indels:	11
Db:	29	Gaps:	5
US-09-889-756A-2 (1-412) x BZ554963 (1-923)			
Qy	19	AlaLeuValLeuSerSerGlyLysGlyAspAlaAlaGlnGlyGlyValProAla	38
Db	695	GCCTTAACACTTCTTCG-----GGGATTTGGCCCCCGGGGGCGCGAGC	648
Qy	39	GlyArgGluAlaProAlaProVal---ValGlyValValThrValHisProGlnThrVal	57
Db	647	GGC-----CCGCAAGCGTTCCGTTCCGAGGTGGTTCGTCG-----GTA	603
Qy	58	AlaIeuThrValGluLeuProGlyArgGluGluSerLeuAlaGlnAlaAspValArgAla	77

DB	602	ACACCTTACGCCAAGTTACACCGCTCGCTGACACCGCTGTCCAGACAGTAAACTGCNCGC	543
Qy	78	glnvalgilylilelleignlysarleupheinglignlysertryValargalagly	97
Db	542	CGCGTCGCGGCTACATTACAGGACGTACCGGTCGGGACGAGCGCGCTGTAGGAAGGCG	483
Qy	98	glnProleuutyrglnilleaspsererthryrglualaaenleugluserlalarGal	117
Db	482	CAGCAACCTGTTCTTCATGACCCGCGCTGTTCAGAGCGCGCAGAGATCCCGCCAGGGA	423
Qy	118	glnleualatrratlaaglnalathrrleualalysalaaaplaaspleualalargtryr	137
Db	422	CGCTTCGCGGAGCGCCAGGACCGCGCTTCCTGCGCGCCAGCAACAGCAGCGACCCG	363
Qy	138	ProleuValalalaaaglnalavalserarglignlutyrgspalalavalalrhalatys	157
Db	362	GTGCTGTATGCGCGGAGTGTGTGCGCCCGGAGCGGCTTCGACAGCGCATTCGCTCGCT	303
Qy	158	Argseralaglnalaglyvallylealalaaaglnalalalelyseralactylleasn	177
Db	302	AATCCAGCAAGGACAGAGTGTGATGCTTCAGGAACGCCCTTGACGCGCGCAATGTGT	243
Qy	178	LeuasnArgserargillethralaProilleserglypheilleglInserlyValser	197
Db	242	ATCGGCTTCGCTCGGTATACACACCGATCGGCTGTGTGGCATATGCAAGTCTT	183
Qy	198	gluglythrleuLeuasnalaglyAspthrthrrValleualathrllearglInthrAsn	217
Db	182	GAGGGTACTACTATGTACCAATGGC---GTGCGCGCTCTTACAGCATTTGTCGGTGTAT	126
Qy	218	PromethylrVal	221
Db	125	CCACTGTACGTG	114
RESULT 22			
LOCUS	BH189931	597 bp	DNA
DEFINITION	ATX060TR ATX0 Arabidopsis thaliana genomic clone ATX060, genomic	linear	GSS 19-OCT-2001
ACCESSION	BH189931		
VERSION	BH189931.1	GI:16302558	
KEYWORDS	GSS.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1 (bases 1 to 597) Bowman, C.L., Van Aken, S.E., Uterback, T.V.,		
AUTHORS	Kaul, S., Town, C.D., Feldblum, T.V. and Fraser, C.M.		
TITLE	Survey sequencing of Arabidopsis thaliana BAC T24G4		
JOURNAL	Unpublished		
COMMENT	Other GSSs: ATX060TF		
	Contact: Chris Town		
	7912 Medical Center Drive, Rockville, MD 20850, USA.		
	Tel: 301-838-3523		
	Fax: 301-838-0208		
	Email: cdtown@tigr.org		
	From Wash. U contig 1142. Caution: the DNA in this BAC may be from		
	some non-Arabidopsis source		
	Seq primer: TR		
	Class: sheared ends.		
FEATURES	Location/Qualifiers		
source	1..597		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/clone="ATX060"		
	/clone_1b="ATX0"		
	/note="Vector: pHO52, Site_1: BstXI, 2-3 kb sheared BAC		
	DNA inserted into pHO52 using BstXI linkers"		

```

BASE COUNT      106 a      192 c      183 g      116 t
ORIGIN

Alignment Scores:
Pred. No.:      1,39e-12      Length:      597
Score:          216.00      Matches:      48
Percent Similarity: 58.54%      Conservative: 24
Best Local Similarity: 39.02%      Mismatches: 47
Query Match:    10,70%      Indels:      4
DB:            28      Gaps:      3

US-09-889-756a-2 (1-412) x BH189931 (1-597)

Qy      273 ValValaAengluSerThrGlyGlnIleThrLeuArgAlaAlaValProAspAspGlnAsn 292
Db      3 GTGGGAGTGAAGACACACCGGATGTGTACGCTGCCCGCGAGTTTCCATCCGACAG 62

Qy      293 IleLeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaValAspAsnAla 312
Db      63 ATCCGTGCTTCCCGGATGTGTGCGCGGTGCGGTGTCAGACCGGTGAAGAGACGTG 122

Qy      313 PheValValProGlnGlnAlaValThrArg-----GlyAlaIlyAspThrValMetIle 330
Db      123 GTCAACCGGACACGCGCGGTGACCCGATGCAGGCGGTGCGGAGCGGTGAGTGC 182

Qy      331 ValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlnIleThr 350
Db      183 ATCGACGATTCACACACGCGGATCCGATGATCCAGATCCAGATCCAGACGCGGTGCGAC 242

Qy      351 AsnTrpIleValIleThrSerGlyLeuIlyAspGlyAspIlyValValIleGlyIlySer 370
Db      243 AAGTGGGTGTGACCTCCGCGGTCAAGGCGCGGAGGAAAGTCATCATGAAAGGCTTC-- 299

Qy      371 IleAlaGlyIleThrGlyAlaIlySerValIleProIlySerGluTrpAlaSerSerGln 390
Db      300 CTGAAGGCTGCTCCCGGAGCG--CCGGTGTTCGAGACCTTCGTCGCCAAGACCGCAG 356

Qy      391 GlnAlaAla 393
Db      357 ACCGCTGCG 365

RESULT 23
BZ561644/c      747 bp      DNA      linear      GSS 17-DEC-2002
LOCUS
DEFINITION      pacs2-164_3364_x1 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION      BZ561644
VERSION      BZ561644.1 GI:27182075
KEYWORDS      GSS.
SOURCE      Pseudomonas aeruginosa
ORGANISM      Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE      1 (bases 1 to 747)
AUTHORS      Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE      Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL      J. Bacteriol., (2002) In press
COMMENT      Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source      1..747
            location/Qualifiers
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="2-164"
            /db_xref="taxon:287"

```

```

BASE COUNT      112 a      238 c      251 g      144 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:      1,89e-12      Length:      747
Score:          216.00      Matches:      56
Percent Similarity: 53.09%      Conservative: 30
Best Local Similarity: 34.57%      Mismatches: 74
Query Match:    10,70%      Indels:      2
DB:            29      Gaps:      2

US-09-889-756a-2 (1-412) x BZ561644 (1-747)

Qy      183 IleThrAlaProIleSerGlyPheIleGlyGlnSerIlyValSerGlyIlyThrLeuLeu 202
Db      582 GTTATCCGCCCCGATTAAACGACCCCGCAGCGGTGTCACAAAGCGCGCTGTGC 523

Qy      203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnPrometTyValAsn 222
Db      522 GCGTAGAATCTGCCGACACTTACCCGCTCAGACAGATCGATCCGATCTACGTGAAC 463

Qy      223 ValThrGlnSerAlaSerGluValMetIlySeruArgGlnIleAlaGlyIlySeru 242
Db      462 TTCCTCCAGCGCGCGGAGAGTCCGCCATGACGAGCGCGCATCCGGAAGCGCAGGTG 403

Qy      243 Leu--AlaAlaAspGlyValIleAlaValIleIlySerPheAspAspGlyThrValTy 261
Db      402 AAGGTGTGCGCGCAAGAGCATCTCCGTGCGCTGCTGCGCGCGCAGCGCAGCATAC 343

Qy      262 ProGluIlySerGlyValLeuLeuPheAlaAspProValValaAngluSerThyGlnIle 281
Db      342 CCGCTGCGCGCGGAGCTGTCTTCCGACTGCGCGGTGCGCGCGCGCAGACCAATC 283

Qy      282 ThrLeuArgAlaAlaValaProAspAspGlnAsnIleLeuMetProGlyIlyLeuTyValArg 301
Db      282 GCCATGCGTCCCTGCTTCGCCAATCCGATCCGATCCGATCCGATCCGATCCGATCCGAT 223

Qy      302 ValLeuMetAspGlnValAlaValaAspAsnAlaPheValValaProGlnGlnAlaValThr 321
Db      222 GTGCGCTCGACGCGCGGTAACCCGACGATCCACCGTCCGCGCGCGCGCGCTGATC 163

Qy      322 ArgGlyIlyAlaIlyAspThrValMet--IleValAsnAlaGlnGlyGlyMetGluProArg 340
Db      162 CGTACCGCCGACGTCGCCGTGTCAAGGTGTCCGACACAGGAGTGAAGAAAGCCCTG 103

Qy      341 GluVal 342
Db      102 GAGGTC 97

RESULT 24
BZ336018      568 bp      DNA      linear      GSS 06-NOV-2002
LOCUS
DEFINITION      hz28c04_g1 WGS-Sbicolorf (JM107 adapted methyl filtered) Sorghum
bicolor genomic clone hz28c04 5', genomic survey sequence.
ACCESSION      BZ336018
VERSION      BZ336018.1 GI:24729117
KEYWORDS      GSS.
SOURCE      Sorghum bicolor
ORGANISM      Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE      1 (bases 1 to 568)
AUTHORS      Rabinowicz,P.D., O'Shaughnessy,A.L., Ballia,V., Dedhia,N.,
Katzenburger,F., King,L., Miller,B., Miller,S., Nascimento,L.,
Zutavern,T., Palmer,L., McCombie,W.R. and Marienissen,R.A.
TITLE      Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
JOURNAL      Unpublished
COMMENT      Contact: W. Richard McCombie

```


Db 381 GAGCCCGCCACCGCCGCCGAGCAAGCGCGGTCCGCCGACCCAGCGCAA 322
 Oy 171 IleuyserrAlaGlyIleAsnleuAsnArgserArgIleThraProIleSerGlyPhe 190
 Db 321 CTGAGCGCGCGCGCTGAACTGAGCTTCACCGGATCAGCGCCCATGACAGCGTGC 262
 Oy 191 IleGlyInserLysValSerGluGlyThreuleuAsnAlaGlyAspThrThrValIleu 210
 Db 261 GTCAAGCCCGCGCGAGGTACCGCGCAACCTGTGCACTCCGGGAG--ACCTGCTC 205
 Oy 211 AlaThrIleArgInThraAsnPromethyrrValaenValThrgInserAlaSerGluVal 230
 Db 204 ACCACCTGCTGTGACGACGACGAGGTCTACCGCTTA-CTTCAGCGCGACGAGCGGTGT 146
 Oy 231 MetLysLeuArgArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
 Db 145 CCTCAAGTACGTGAGCTGCGCGCCGACGCGCGGTGCGACACAGCGCCAGACGCCCGGT 86
 Oy 251 ValGlyIleLysPheAspAspGlyThrValTyrrProGluysGlyArgLeuLeuPheAla 270
 Db 85 GTGGGG----- 80

Oy 271 AsproValValaenGlu 276
 Db 79 GATCCACTAGTTCTAGAG 62

RESULT 26

B2551287/c

LOCUS B2551287 809 bp DNA linear GSS 17-DEC-2002
 DEFINITION pasci-60_3076.y2 pasci-60 Pseudomonas aeruginosa genomic clone
 ACCESSION B2551287
 VERSION B2551287
 KEYWORDS GSS

SOURCE

ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.

REFERENCE

AUTHORS

1 (bases 1 to 809)
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE

Whole-Genome-Sequence Variation among multiple isolates of
 Pseudomonas aeruginosa library

JOURNAL

J. Bacteriol., (2002) In press

COMMENT

Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu

Class: shotgun.
 Location/Qualifiers

FEATURES

source

1..809
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pasci-60_3076"
 /clone_11b="pasci-60"
 /note="Clinical isolate 1-60 whole genomic shotgun
 library."

BASE COUNT 138 a 241 c 261 g 169 t

ORIGIN

Alignment Scores:

Pred. No.: 4.37e-10 Length: 809
 Score: 194.00 Matches: 60
 Percent Similarity: 49.73% Conservative: 31
 Best Local Similarity: 32.79% Mismatches: 58
 Query Match: 9.61% Indels: 34
 Gaps: 7

US-09-889-756a-2 (1-412) x B2551287 (1-809)

Oy 42 AlaProAlaProValaGlyValIleValThraHisProGlnThraValaLeuThraVal 61

Db 564 GCACCAAAACCGCCGATCAGGTG-----ACCGCAGCTTGGCCGAA 523

Oy 62 GlnLeuPro-----GlyArgLeuGlnSerLeuArgThrAla 73

Db 522 AAGGTCCTTGGCAGACCGCTCCAGCATCCGCGACGCTCAAGCATTCACAGGCGTG 463

Oy 74 AspaValArgAlaGlnValaGlyIleIleGlnLysArgLeuPheGlnGluSerTyr 93

Db 462 ACCCTACCGCGGAGGTCTCCGACGCTACGACGACGATCTCTTCCGGGACCCAG 403

Oy 94 ValArgAlaGlyGlnProLeuTyrrGlnIleAspSerSerThrTyrrGluAlaAsnLeuGlu 113

Db 402 GTGAAGCTGACACCAACCGCTGATCAGTTGGAAGACACGTCGAGAACCACTTCGCG 343

Oy 114 SerAlaArgAlaGlnLeuAlaThraAlaGlnAlaThraLeuAlaValaAspAlaAspLeu 133

Db 342 -----ACTGCCAGGCCGATCTCCGCTGCGCCAGGCGGAGTAC 304

Oy 134 AlaArgTyrrLysProLeuValaAlaAlaGluAlaValaSerArgGlnTyrrAsp----- 151

Db 303 CAGCGCGCGCGGACGACGATCGGACGAGGCGCATCTCGAAMACGATTCGATCGTCTC 244

Oy 152 AlaAlaValThraAlaLysArgSerAlaGluAlaGlyValLysAlaAlaAlaAlaIle 171

Db 243 GCCCGCAGTGGGCGGACGACGACGACCGCTC-----GCCGAGCTG 202

Oy 172 LysSerAlaGlyIleAsnleuAsnArgserArgIleThraAlaProIleSerGlyPheIle 191

Db 201 AAGCGCGG-----CTGGCAGAGAGCGCGTGTGCGCGCTTCGCGGAGACCATC 151

Oy 192 GlyInserLysValSerGluGlyThreuleuAsnAlaGlyAspThrThrValleuAla 211

Db 150 GGCAATCCGCGGAGGTGAGAGCTGCGGACTGCTCGCCGCGG-----ACCGCATGCGC 97

Oy 212 ThrIleArg 214

Db 96 ACCATCAG 88

RESULT 27

CA554488

LOCUS

CA554488 595 bp mRNA linear EST 19-NOV-2002
 DEFINITION C0886C11-5N NIA Mouse Blastocyst cDNA library (Long) Mus musculus

CDNA clone NIA:C0886C11 IMAGE:30033250 5', mRNA sequence.

ACCESSION

CA554488

VERSION

CA554488.1

KEYWORDS

EST.

SOURCE

Mus musculus

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Other ESTs: C0886C11-3

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333. Cassell Drive, Suite 4000, Baltimore, MD 21224-6620, USA

Email: cdna@igsun.grb.nia.nih.gov

Plate: C0886 row: C column: 11

Seq primer: M13 Reverse

High quality sequence stop: 595

POLYA=No.

FEATURES

source

Location/Qualifiers

1..595

/organism="Mus musculus"

/mol_type="mRNA"


```

/strain="CS7BL/6J"
/db_xref="nblast:C0886C11-5N"
/db_xref="taxon:10090"
/clone="NIA:C0886C11 IMAGE:30033250"
/tissue_type="Blastocyst"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI. Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsn.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 20 Blastocysts. Double-stranded
cDNAs were synthesized with an oligo(dT) primer
[Invitrogen:
5'-pGACTGATGTTCTAGATCCGAGCGCCGCCCTTTTCTTTT-3'] from
0.2 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker lp-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.2 kb. The library was constructed
by Yulan Piao (NIA)."

```

BASE COUNT 199 a 116 c 124 g 156 t

ORIGIN

Alignment Scores:

Pred. No.: 8.54e-10 Length: 595
Score: 189.50 Matches: 59
Percent Similarity: 49.13% Conservative: 54
Best Local Similarity: 25.65% Mismatches: 82
Query Match: 9.39% Indels: 35
DB: 14 Gaps: 8

US-09-889-756a-2 (1-412) x CA554488 (1-595)

```

QY 78 GlnValGlyGlyIleGlnLysArgLeuPheGlnGlySerTyrValArgAlaGly 97
Db 3 GAAGTAAGTGGCCGTTTAATCTTCAATCTCAAGAGGAGGAAATGATCTCAAGGA 62
QY 98 GlnProLeuTyrGlnIleAspSerSerThrTyrGlnLysLeuGlnSerAlaArgAla 117
Db 63 ACGGTGCTGCGAGGATCAATGATGATGCTACACGCAATATGAGAAATTCGGTGT 122
QY 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLys 137
Db 123 CAATTGGATCTGCACAA-----AAAATGAAAGCTCGACTGAAAA 161
QY 138 ProLeuValAlaAlaGlnAlaValSerArgGlnGlyTyrAspAlaAlaThrAlaLys 157
Db 162 AAATGCTGATATCAATGATATCAATGATGATGATATATATGCGCACTCAATCAGGTG 221
QY 158 ArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLysSerAlaGlyLeuAsn 177
Db 222 AATGCTTAAAGAGATCTGATATCTCAATGATGATGATGATGATGATGATGATGATGAT 263
QY 178 LeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSer 197
Db 264 ---GATAAACAGGTGTTAAAGACCTTTCACCGGTATATAGATTGAGAAATGATCAGC 320
QY 198 GlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsn 217
Db 321 CCGGTCCTTATGTAAACA-----CCAGCTACCAATTTCTTCAGCTTACAAACCAAGAT 374
QY 218 ProMetTyrValAsnValThr-----GlnSerAlaSerGlnValMetLysLeuArgArg 235

```

```

Db 375 AAAGTAAGATCATTTTACTGTTCCGAATGATATCGACGTT----- 419
QY 236 GlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPhe 255
Db 420 ---ATTGTTAAAGGAAAA-----TCAGTCAATGTAAACC 452
QY 256 AspAspGlyThrValTyrProGlnLysGlyArgLeuLeuPheAlaAspProValAlaAsn 275
Db 453 AATGCGATGGGTCTGTT---CCGAAGCCACCATCTTCTGATGACACAGATGAT 509
QY 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMet 295
Db 510 GCTACTACCCGTTACCTTAAGTAAGACGACGATATG-----GATGAGCCCAATATTAAT 563
QY 296 ProGlyLeuTyrValArgValLeuMetAsp 305
Db 564 CCGGTCCTTTTGTAAAGTATTAATGAT 593

```

RESULT 28
B2558130
LOCUS B2558130 756 bp DNA linear GSS 17-DEC-2002
DEFINITION pacsl-60_808.s1 pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION pacsl-60_808, genomic survey sequence.
VERSION B2558130
KEYWORDS B2558130.1 GI:27171967
SOURCE GSS.
ORGANISM Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 756)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence Variation among Multiple Isolates of
Pseudomonas aeruginosa Library
J. Bacteriol., (2002) In press
CONTACT: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
CLASS: shotgun.
FEATURES
source location/Qualifiers
1..756
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_808"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 whole genomic shotgun
library."

BASE COUNT 140 a 263 c 223 g 130 t

ORIGIN

Alignment Scores:
Pred. No.: 1.18e-09 Length: 756
Score: 189.50 Matches: 67
Percent Similarity: 48.18% Conservative: 39
Best Local Similarity: 30.45% Mismatches: 77
Query Match: 9.39% Indels: 37
DB: 29 Gaps: 9

US-09-889-756a-2 (1-412) x B2558130 (1-756)

```

QY 14 LeuAlaAlaValAlaLeuValLeuSerSerCysGly---LysGlyGlyAspAlaAla 32
Db 149 CTCGCGCGGTATGCGCGTGGTGGCGGATTCGCGCGGCTACAAAGCTTCAATCCGCT 208
QY 33 GlnGlyGlnProLeuAlaGlnAlaProAlaProValValGlyValValThrVal 52

```

```

Db 209 CAG-----CAGATCGCCCTTTTCAGCCGACCGAAGCCGCGATCAGCGTG----- 253
Qy 53 HisProGlnThrValAlaLeuThrValGluLeuPro----- 64
Db 254 -----ACCGCGCGCTGGCCGAAAGCGTCCGTGGCAGAGCCGCTCCGACCGATC 304
Qy 65 GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleLeuGln 84
Db 305 GCGACGCTCAAGCATTCACGAGCGCTGACCTCACCCTCGAAGTCTCCGACCGGTACG 364
Qy 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAsp 104
Db 365 GACGTAAGTCTTCCTTCGCGCGACGACGAGTGAAGCTGACCAACCGCTGATCAGATTGAA 424
Qy 105 SerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 425 AGCGACGCTGAGAAAGCCACCTGCGC-----ACTGCGCAGGCGC 463
Qy 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAla 144
Db 464 GATCTCGGCGCTGCGCAGGCGCGAGTACGAGCGCGCGCGAATGATCGGACGACAGGCC 523
Qy 145 ValSerArgGlnGluTyrAsp-----AlaAlaValThrAlaLysArgSerAlaGlnAla 162
Db 524 ATCTCGAAAGCGAATTCGATGCTCGCGCGCAGTGGCGCAAGACGACGCCACCGCTC 583
Qy 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 584 -----GCCGAGCTGAAAGCGCGCG-----CTGCGCAGAAAGCGC 616
Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202
Db 617 GTGCTCGGCGCTTCCGCGCGGACCATCCGCACTCCCGCAGGTGAGCTGCGACATCAGTC 676
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db 677 TCGCCCGGG-----ACGCGCATCGCAACTTGACAGACCTTCCACCTGCTCTGAT 730

RESULT 29
LOCUS B2340887 697 bp DNA linear GSS 06-NOV-2002
DEFINITION B2340887 B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION B2340887
VERSION B2340887.1 GI:13775936
KEYWORDS GSS
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE Tomkins,J.P., Wood,T.C., Stacey,M.G., Itoh,T.T., Judd,A., Goicoechea
AUTHORS 'J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 711.
Location/Qualifiers
1..775
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"

```

```

BASE COUNT 168 a 250 c 215 g 139 t
ORIGIN
Alignment Scores:
Pred. No.: 1 76e-09 Length: 775
Score: 188.00 Matches: 58
Percent Similarity: 46.24% Conservative: 28
Best Local Similarity: 31.18% Mismatches: 89
Query Match: 9.31% Indels: 11
DB: 28 Gaps: 3
US-09-889-756a-2 (1-412) x B2340887 (1-775)
Qy 22 LeuSerSerGlyGlyGlyGlyAlaAlaGlnGlnGlyGlnProAlaGlyArgGlu 41
Db 215 CTGCGCGAATGCGCGAAGAAAGCCGACGCGCAAGAA-----AGGTCC 256
Qy 42 AlaPro-AlaProValValGlyValValThrValHisProGlnThrValAlaLeuThrVal 61
Db 257 GCGCGCGCGCGCGCGCTCCGTCGAGCGTGGCGCGCGTCCAGAGGTGATTCGCCGCT 316
Qy 61 GlyLeuPro-----GlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnAla 79
Db 317 CTACTTACCGGTCTCGGACAGGTTCAGGTTTCAACACCTCCAGGTCCGCCCGCGT 376
Qy 79 GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnPro 99
Db 377 TGACGCGCAGATCGACAGATCGCTTTACGAGAGCGCAGATCGTCAAGAGCGCGAGCT 436
Qy 99 OleuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLe 119
Db 437 TCTGTCTCGATCGATCCCGCTCCCTATCAAGCGCGCTCGACGCGCAAGGCGCAAGAA 496
Qy 119 ValAlaThrAlaGlnAlaThrAlaLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysPro 139
Db 497 GCGCGACGACGAGCAATTCGCGCAAGCGCAATCTGAGCTCCAGCGCGCATGAAGCT 556
Qy 139 ValAlaAlaGlnAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSe 159
Db 557 CGCGGAATTCGCGACCGCGCGAGCGTGCACACCGACCGCTCCACTGCGCGCGAGTTAC 616
Qy 159 ValAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAs 179
Db 617 CGNCCAGATCGCG-----GCCGACAAAGCGCGGATCTCCAAATCCGAAACCACTTTGA 670
Qy 179 nargSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlu 199
Db 671 CTACACCAAGTCAAGGAGGCGGATCCAGCGGGGTGCGCGGCGCGCAGGTGACACCGTGC 730
Qy 199 YThrLeuLeuAsnAla 204
Db 731 GAACATGTCTCAAGCC 746

RESULT 30
LOCUS B2340887 697 bp DNA linear GSS 06-NOV-2002
DEFINITION B2340887 B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION B2340887
VERSION B2340887.1 GI:24739302
KEYWORDS GSS
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE Rabinowitz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
1 (bases 1 to 697)

```



```

Db      579  CGGCTACGTCGAGGTGGCCNTGACGCGCGGTGACCGAAGCGGATCACCGTCCGCG 520
Oy      317  nglAlaValThrArgGlyAlaValAspThrValMet---lIleValaAlaGlnGly 336
Db      519  CGAGCGGTGATCCGTACCGCCAGTCCGCGGTGATGATGATGATGATGATGATGATGATGAT 460
Oy      336  yMetGluProArgGluValThrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
Db      459  GGTGGAACAGTGAAGTCCCGCGCGACACCTGAGGCGGTGATGATGATGATGATGATGATGAT 400
Oy      356  rGlyLeuIleuAspGlyAspIleValValAlaGlnGlyIleSerIleAlaGlyIleThrG 376
Db      399  CGGGCTCAAGGCGCGCGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
Oy      376  yAlaIleuValThrProArgGluThrAlaSerSerIleuGlnGlnAlaAlaAlaProG 396
Db      356  -----CAGCATGCGCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 316
Oy      396  nSerGlyValAlaGlnThrAlaSerGluAlaValThrAlaSerGluAla 411
Db      315  GCCGCGCAGCGCGCGCGCGCTACCGCTGCGCGGTGCGCGCG 270

RESULT 32
LOCUS   AZ302989/c 472 bp DNA linear GSS 06-MAR-2001
DEFINITION  GSSBnu1931 Brucella abortus random genomic library Brucella
              melitensis biovar Abortus genomic clone U01931, genomic survey
              sequence.
ACCESSION  AZ302989
VERSION    AZ302989.1 GI:10129200
KEYWORDS   GSS.
SOURCE     Brucella melitensis biovar Abortus (Brucella abortus)
ORGANISM   Brucella melitensis biovar Abortus
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Brucellaceae; Brucella.
REFERENCE  1 (bases 1 to 472)
            Sanchez,D.O., Zandemmi,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,
            Faccio,P., Diaz,G., Lanzavecchia,S., Agnello,F., Frasch,A.C.C.,
            Anderson,S.G.E., Rosetti,O.L., Grau,O. and Ugaldede,R.A.,
            Gene discovery through genomic sequencing of Brucella abortus
            Infect. Immun. 69 (2), 865-868 (2001)
            21101034
            11159979
COMMENT     Contact: Siv Andersson
            Small Genomes Sequencing Group
            Department of Molecular Evolution, Uppsala University
            Nobdvägen 18C, S-752 36, Uppsala, Sweden
            Tel: 46-18-471-4379
            Fax: 46-18-471-6404
            Email: Siv.Andersson@bc.uu.se
            Sequences were basecalled with phred and vector was masked with
            crossmatch (see http://genome.washington.edu). Sequences were then
            trimmed from both ends to remove low quality bases and masked
            vector.
            Class: shotgun.
FEATURES             Location/Qualifiers
     source           1..472
                     /organism="Brucella melitensis biovar Abortus"
                     /mol_type="genomic DNA"
                     /strain="2308"
                     /db_xref="taxon:235"
                     /clone="U01931"
                     /clone_1id="Brucella abortus random genomic library"
                     /note="Vector: modified M13"

BASE COUNT      88 a 131 c 140 g 113 t
ORIGIN
Alignment Scores:
Pred. No.:      5.52e-09      Length: 472
Score:          180.50      Matches: 48
Percent Similarity: 51.01%      Conservative: 28
Best Local Similarity: 32.21%      Mismatches: 62

```

```

Query Match:      8.94%      Indels: 11
DB:               28      Gaps: 3
US-09-889-756a-2 (1-412) x AZ302989 (1-472)
Oy      249  lIleAlaValGlyIleLeuPheAspAspGlyThrValIleProGlnGlyArgLeu 268
Db      466  GTAGCCGTGACCGCTGACGCTCCGACGCGATGATATGATGATGATGATGATGATGATGAT 407
Oy      269  PheAlaAspProValValaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 288
Db      406  TTATCCGGATCAAGATTATATCCCATATCCGATACGTTATGATGATGATGATGATGATGAT 347
Oy      289  AsnAspGlnAsnIleLeuMetProGlyLeuIleValArgValLeuMetAspGlnValAla 308
Db      346  AATCCGAACATGTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 287
Oy      309  ValAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgIleAlaValAspThr 327
Db      286  CCGGTGAACCCCTGCTGCTGCGCAAGCAATCATGATGATGATGATGATGATGATGATGATGAT 227
Oy      328  ValMetIleValaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 347
Db      226  GTTCTGCGCGTGGTGAAGATTAACAGATCATCCAGCGCATGATGATGATGATGATGATGATGAT 167
Oy      348  GlnGlyThrAsnThrIleValThrSerGlyLeuIleuAspGlyAspValValAlaGlu 367
Db      166  TTTCGACAGCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 107
Oy      368  GlyIleSerIleAlaGlyIleThrGlyAlaValValValValValValValValValValVal 380
Db      106  GGCTTCAG-----CCCATCCGCGCGCGCGAGAGGTGATGATGATGATGATGATGATGATGAT 53
Oy      381  ---ThrProArgGluThrAlaSerSer 388
Db      52  ACCAGCGCGCGCGCGCAAGAGCATGTCG 26

RESULT 33
LOCUS   BH247218/c 450 bp DNA linear GSS 26-NOV-2001
DEFINITION  BOGAB20TR BOGA Brassica oleracea genomic clone BOGAB20, genomic
              survey sequence.
ACCESSION  BH247218
VERSION    BH247218.1 GI:17067767
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
            1 (bases 1 to 450)
REFERENCE  Town,C.D., Van Aken,S., Uteirback,T., Koo,H. and Fraser,C.M.
            Whole genome shotgun sequencing of Brassica oleracea
            unpublished
            Other GSSs: BOGAB20TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..450
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGAB20"
                     /clone_1id="BOGA"
                     /note="Vector: pHD01, Site_1: BseXI; 2-3 kb sheared

```


Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
pac62-164_3895									
BZ562720									
BZ562720.1	GT:127185220								
GS.									
Pseudomonas aeruginosa									
Pseudomonas aeruginosa									
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.									
1 (bases 1 to 847)									
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.									
Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library									
J. Bacteriol., (2002) In press									
Contact: Chris K. Raymond									
Genome Center									
University of Washington									
Box 352145, Seattle, WA 98105-2145, USA									
Tel: 2062216954									
Fax: 2066857244									
Email: ckraymond@u.washington.edu									
Class: shotgun.									
Location/Qualifiers									
1. 847									
/organism="Pseudomonas aeruginosa"									
/mol_type="genomic DNA"									
/strain="2-164"									
/db_xref="taxon:287"									
/clone="pac62-164_3895"									
/clone_lib="pac62-164"									
/note="clinical isolate 2-164 Whole genomic shotgun library."									
BASE COUNT	196 a	269 c	236 g	146 t					
ORIGIN									
Alignment Scores:									
Pred. No.:	6.7e-08	Length:	847						
Score:	173.50	Matches:	61						
Percent Similarity:	43.06%	Conservative:	32						
Best Local Similarity:	28.24%	Mismatches:	86						
Query Match:	8.59%	Indels:	37						
DB:	29	Gaps:	6						
US-09-889-756A-2 (1-412) x BZ562720 (1-847)									
Qy	96	AlaGlyIcInProLeuTyrgInIleAspSerSerThrTyrgInAlaAsnLeuGInSerAla	115						
Db	1	AGCGGCGCTTCCTCTACTAGTACGTAAGAACTGACGTCACCGCG--GTGGCGGCTCT	57						
Qy	116	ArgAlaGInLeuAlaThrAlaGInAlaThrLeuAlaThrLeuAlaAlaAspAlaAspLeuAlaArg	135						
Db	58	CTAGAACTAGTGGATTCCTCCCTCGCGGCACTGAAAGACCGCGAGATGACCTTCAGCGC	117						
Qy	136	TyrlsProLeuValAlaAlaGInAlaValAspArgInGluTyrgAspAlaAlaValThr	155						
Db	118	TACAGAGGCGCTGTAACGCGCAGACTCGATCGCGAAGACCTCGATATACCAGAAAGCC	177						
Qy	156	AlaIysArgSerAlaGInAlaGInValIysAlaAlaGInAlaAlaIleIysSerAlaGly	175						
Db	178	CAGGTCCGCGACATTGACGACCACTGATCAACACAGGCGCAGGTGACGACGACCGCGC	237						
Qy	176	IleAsnLeuAsnArgSerAlaThrAlaProIleSerGluPheIleGlyGInSerIys	195						
Db	238	CTCAACCTGACCTTACCGACGAGGTCCGCGACCGATTTCCGGCGGCTTCGCGCGAG	297						
Qy	196	ValSerGluGlyThrLeuLeuAsnAlaGInIysAspThrThrValLeuAlaThrIleArgIn	215						
Db	298	GTGACATCGGCAACCTGTCACACGACGCGATACACAGCGCGCTGTGTGATCACCGAG	357						
Qy	216	ThrAsnProMetIyValAlaValThrInSerAlaSerGluValMetIysLeuArgArg	235						
Db	358	GTCAGAGCCATC-----TCGGTGT							

Accession	Source	Strain	Length	Score	Similarity	Percent Similarity	Best Local Similarity	Query Match	DB
236	GenBank	Glut1lealagluclylyLeuLeuLeu1la1aspGlyVal1lealValc1lyleuyspne	255						
394	CAG	-----	396						
256	AspaSpiglythrValTyr	-----ProGluLygGlyArgLeuLeuPhe	270						
397	CAGATCGGCACCGCTGTCGAGCATGACGCGCCCGCAGCATGACGCTCACCCGCTG	456							
271	AspProValValaenGluSerThr	-----GlyGln1leThrLeuArg1la1aVala	287						
457	GACCCGAACCAAGAGAGGTTCTTCCGGAAGAAGGACCCCTACCAACCCCTTGACACAAAC	516							
288	ProAsnaSpGlnAsn1leLeuMetProGlyLeuTyrValArgValLeu	303							
517	CCA---GATTGGAACACACCTACTCCGGGGAACGAGTCAAAAGCTT	561							
RESULT 36	BZ549534/C	793 bp DNA linear	GSS 17-DEC-2000						
LOCUS	pacsl1-60_2148, genomic survey sequence.								
DEFINITION	BZ549534.1 GI:27153115								
ACCESSION	BZ549534								
VERSION	BZ549534.1								
KEYWORDS	GSS.								
SOURCE	Pseudomonas aeruginosa								
ORGANISM	Pseudomonas aeruginosa								
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.								
AUTHORS	1 (bases 1 to 793) Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.								
TITLE	Whole-Genome-Sequence Variation among Multiple Isolates of Pseudomonas aeruginosa Library J. Bacteriol., (2002) In press								
JOURNAL	Contact: Chris K. Raymond								
COMMENT	Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062316954 Fax: 206657244 Email: ckraymond@u.washington.edu Class: Shotgun.								
FEATURES	Location/Qualifiers								
source	1..793 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="1-60" /db_xref="taxon:287" /clone="pacsl1-60_2148" /clone_1ib="pacsl1-60" /note="Clinical isolate 1-60 whole genomic shotgun library"								
BASE COUNT	146 a 245 c 250 g 151 t	1 others							
ORIGIN									
Alignment Scores:									
Pred. No.:	7.8e-08	Length:	793						
Score:	172.50	Matches:	51						
Percent Similarity:	50.00%	Conservative:	18						
Best Local Similarity:	36.96%	Mismatches:	64						
Query Match:	8.54%	Indels:	5						
DB:	29	Gaps:	2						
US-09-889-756A-2 (1-412) x BZ549534 (1-793)									
QY	16 A1a1aVala1aLeuValLeuSerSerCySgLySgLyGlyAspAla1aGlnGlyGly	35							
DB	437 GCGGCGCCGCGCTGCGCGCTGCTCGACGCGGCAACTCGGT	387							
QY	36 GlnProAlaGlyArg-----GlnAlaProAlaProValValGlyValValThrValHis	53							
DB	386 AATCCGCGCGCGCGCTCCCAAGGCGCAACGCGCTCACCGTGGCGTGGCCAGGCGTGGAG	327							

QY 54 ProGlnThrValAlaLeuThrValGluLeuProGluLeuGluSerLeuArgThrAla 73
 Db 326 CAGGCGCAGCTGGCGGTCATTTTCACACCGCTTCCACCGCTTCCACACCGG 267
 QY 74 AspValAlaArgAlaGlnValGlyGlyLeuLeuGlnValArgLeuPheGlnGluGlySerTyr 93
 Db 266 AACGTCACAGCCGGGCTACCGGAGCTGTCTCAAGGTGCTTCCAGAGGGGCGAGAG 207
 QY 94 ValArgAlaGlnGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlu 113
 Db 206 GTCAAGGCGCGGACCTGCTGGCGGTGTGACACCGCGACCTTCAAGCGCGGCTGACC 147
 QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAspAlaAspLeu 133
 Db 146 CAGCGCCAGGCGCAGCTGATACAGATCCAGCGCAGCACTGAAGAACAACAGATGACGTG 87
 QY 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnGluTyrAsp 151
 Db 86 CAGCGGACACAGGCTGTGTATGCGCTGACGTCTCCGGGCGAGCCGGGAT 33

RESULT 37
 BZ638338 876 bp DNA linear GSS 29-JAN-2003
 LOCUS BZ638338/c
 DEFINITION OGCGN80TC_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBma0141M16,
 genomic survey sequence.
 ACCESSION BZ638338
 VERSION BZ638338.1 GI:28090523
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays

REFERENCE
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
 'A., Fraser,C.M., Budiman,W.A., Bedell,J.A., Kohlfing,T., Citek
 'R.W., Nuberger,A., Robbins,D. and Lakey,N.
 Consortium for Maize Genomics
 TITLE Unpublished
 JOURNAL Other GSSs: OGCGN80TM
 COMMENT Contact: Cathy Whitelaw

FEATURES
 Source
 1. 876
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBma0141M16"
 /clone_1kb="ZM 0.7_1.5 KB"
 /note="Vector: pBCKS-; Site 1: HincII; 0.7-1.5 kb
 methylation filtered genomic DNA library"

BASE COUNT 114 a 287 c 326 g 149 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.45e-07 Length: 876
 Score: 170.50 Matches: 82
 Percent Similarity: 40.58% Conservative: 43
 Best Local Similarity: 26.62% Mismatches: 120
 Query Match: 8.44% Indels: 63
 DB: 29 Gaps: 12

US-09-889-756a-2 (1-412) x BZ638338 (1-876)
 QY 97 GlyGlnProLeuTyrGlnIleAspSerSerThrTyr----- 108

Db 856 GCGCAGCCAAAGCTCTTCATCAAGACGAGCATTTGTCGATGCCGCTGCTTCT 797
 QY 109 ---GlnAlaAsnLeu-GluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAl 127
 Db 796 CCGCGCGCGCGGCTACCAAGACCGCGAGCGCTGTGCGCGCGAG-----CG 746
 QY 127 AluValaAspAlaAspLeuAlaArgTyrProLeuValAla-AlaGlnAlaValaSer 147
 Db 745 CAAGCGCGACCGCAGACGAAATATATAGACCGCGCGCGCGCGGATATCAGC 686
 QY 147 rglngluTyrAsp-----AlaAlaValThrAlaValAspSerAlaGlnAlaGlyVal 165
 Db 685 AGCGCAGGCGGACCTCGCGCGCGCGAGCGACCGCGCGCGCGCGCGCTGCTG 626
 QY 165 ysaAlaGlnAlaAlaIle-----LysSerAlaGlyIleAsnLeuAsnArg- 180
 Db 625 GCGCGCGCGCGCAGACGCTGCGATCTGTGGGAGATGCGCAGGCG--GAGATCAACCGC 569
 QY 181 -----SerArgIleThrAlaProIles 188
 Db 568 TGGAGATGTCGCGAAGTCTCCGACATCCAGACATACACGCTGACGACCGATCA 509
 QY 188 erGlyPheIleGlnGlnSerLysValSerGlnGlyThrLeuAlaGlnAlaGlyAspThr 208
 Db 508 GCGCGCTGCTGCGCTGCGCGGATGTCGCGCGGCGAATATGTGTCAGGCGCGGACA 449
 QY 208 hrValLeuAlaThrIleArgGlnThrAspProMetTyrValaAsnValThrGlnSerAla 228
 Db 448 AGCGGTGATGACATC-----ACCGATCGTCCGCGCTA----- 414
 QY 228 erGluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuAlaAlaAspGly 248
 Db 413 -----TGCCTGTCGCGCGCAGGTGGCGAGCGCAT----- 384
 QY 248 alIleAlaValGlyIleLysPheAspAspGly-----ThrValTyrProGluLys 265
 Db 383 -----GCCGAGATGTCGCGTGGCGGATCCGTCAGAGTACGACCGCGCTGCGG 329
 QY 265 lYArgLeuLeuPheAla-----AspProValValaGlnGluSerThrGlyG 280
 Db 328 GCCGCTCTTCCACGCGGACGATGATGTCGCGCGCGCGCTGACCGCGACGAC 289
 QY 280 lInLeuThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyr 300
 Db 268 GCGTGGCGGCTGCGCGCTGATCCCAACCGCAGACGCGCTGAAGCCCAATGTCG 209
 QY 300 alArgValLeuMetAspGlnValAlaValaAspAsnAlaPheValValProGlnGlnAla 320
 Db 208 CCAGCTTCCGATCAAGCATCTGAACGCGGCGAAGCGATCCGCGCGCGCGCG 149
 QY 320 alThr---ArgGlyAlaLysAspThrValMetIleValaAsnAlaGlnGlyMetGluP 339
 Db 148 TCATTCATGAAGCGACACCGCGCGCTGTGATCTGATGCGCGCGACGCGCTGCTGTCG 89
 QY 339 roArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeu 359
 Db 88 CCGCGCAGCTGACGAGGCGGCGACTCCGCCAATGCTGTGTCACCATCACCGAGGCTGA 29
 QY 359 ysaAspGlyAspLysValVal 365
 Db 28 AGACCGCGCAGAAAGATGCTC 9

RESULT 38
 BP416737 469 bp mRNA linear EST 28-NOV-2000
 LOCUS BP416737
 DEFINITION UI-R-CAO-bku-g-12-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone
 UI-R-CAO-bku-g-12-0-UI 3', mRNA sequence.
 ACCESSION BP416737
 VERSION BP416737.1 GI:11404688
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

Db 191 GAAGGCGAGATGTCGAGGCGGCGAGTTGCTGGCGACCATCGATGACCGAGCGAGTTC 250
Qy 110 AAlaenleuGluserAlaArgAlaGlnleuAlaThrAlaGlnAlaThrleuAlaAla 129
Db 251 CTCGGCGTGGACGACGCGCGCTCCAGCGGAGTAAACAGGCCACCTGAAATCCGAC 310
Qy 130 AspAlaAspLeuAlaArgTyrLeuProleuValAlaAlaGlnAlaValSerArgGln 149
Db 311 GAGCAGACCTGCAACGCTACCGAGCTGTTGGCGGCGGTGGTAACTCTACTG 370
Qy 150 TyrAspAlaAlaValThrAlaLysSerSerAlaGlnAlaGlyValAlaAlaGlnAla 169
Db 371 TTGATTCAGCAACAGCGCGCTCCAGACGCTGCGCGACCCCTTAATCCACAGATCC 430
Qy 170 AlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGly 189
Db 431 ATCATCAACCCCGACGACGCTGCGCTGTAATCACTCGGATCACTCGTCCGTCGCGGT 490
Qy 190 PheIleGlyGlnSerLysValSerGlnGlyThrleuLeuAsnAlaGlyAspThrVal 209
Db 491 AGGGTCGGTATCCGAAACGTCATGTCGAAACGCGAGCGGTCGTCGACAGCTTCTCT 550
Qy 210 LeuAlaThrIleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGlu 229
Db 551 CTGTTCAAGTGGACCAAGATCGATCGATGAGTCTTCTCTCACTGCAACAGGAAAG 610
Qy 230 ValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIle 249
Db 611 TTGCGCCAGTTGC--AGGCACCTGCTCGCGCGGGAAGCGGAGTGCACCCCTACAGTCA 668
Qy 250 AlaValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArgLeuLeuPhe 269
Db 669 ACAGTTGAGGCTCCCTCCGACAGAACGCCGTTGCTTGCACCATCG3----- 713
Qy 270 AlaAspProValValAsnGlnSerThrGlyGlnIleThrleuArgAlaAlaValPro--- 288
Db 714 ---CAACCAATCGCGCTTCAACCGGA-----ACCTTCGCGGAGCGCGCCCTTC 764
Qy 289 -----AsnAspGlnAsnIleLeuMetProGlyLeu 298
Db 765 CTTATATTAACNCGCAGGCGCGCTTCTTGGCCCGGTCTA 806

RESULT 40
BZ578763 1228 bp DNA linear GSS 17-DEC-2002
LOCUS msh2.5991.y2 msh Pseudomonas aeruginosa genomic clone msh2_5991,
DEFINITION Genomic survey sequence.
ACCESSION BZ578763
VERSION BZ578763.1 GI:27213824
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa; Gammaproteobacteria; Pseudomonadales;
Bacteria; Proteobacteria; Pseudomonas.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL U. Bacteriol., (2002) In press
COMMENT "Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1..1228
location/Qualifiers
1..1228
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"

```

```

/db_xref="taxon:287"
/clone="msh2_5991"
/clone_id="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT 226 a 401 c 35t g 242 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1,26e-06 Length: 1228
Score: 163.50 Matches: 64
Percent Similarity: 43.75% Conservative: 41
Best Local Similarity: 26.67% Mismatches: 100
Query Match: 8.10% Indels: 36
DB: 29 Gaps: 6
US-09-889-756a-2 (1-412) x BZ578763 (1-1228)
Qy 8 AlaMetArgAlaAlaAlaAlaAlaAla-----AlaValAlaLeuValLeuSerSerCys 25
Db 206 GCCATGAAGCGGTTTCTCTCCGCGCGCTGTCGCTTCCGCTCTCTCCGCGCTGC 265
Qy 26 GlyLysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaPro 45
Db 266 GGC-----GACGAGCGCGCGCGCACCG 289
Qy 46 ValValGlyValValThrValHisProGlnThrVal-----AlaLeu 59
Db 290 CCGCGCGCGGCTGCTGACGCGTGAACAGACCTGAAAGAACGACGCTCGCGCTTC 349
Qy 60 ThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 350 GCCGAGCATTCACGAGCGCGCTACGAGAGCGTG-----CTCGGCTTCCGAC 397
Qy 80 GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnPro 99
Db 398 AACGAGAGATTCGCTCCGCTGCTGACGCTGATCACTTCCGCGCAAGGCGCGCTG 457
Qy 100 LeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
Db 458 CTGGCGACCTTCGACCCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 517
Qy 120 AlaThrAlaGlnAlaThrleuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeu 139
Db 518 GCCAGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 577
Qy 140 ValAlaAlaGlnAlaValSerArgGlnGlnTyrAspAlaAlaValThrAlaLysArgSer 159
Db 578 TTGCTCGCAGCGCTCACCGCGCGCGCGCTGAGCATCGCGGACCGCTCGAAGACC 637
Qy 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
Db 638 AGCCAGCGCGCTTCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 697
Qy 180 ArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGly 199
Db 698 TACACGCGCGCTGTGACCATTTCCAGCGCGCTATT----- 723
Qy 200 ThrleuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAspProMet 219
Db 734 ACCACCTGTCACCGCGCGCGCGCATGTGTTCCAGCGCGCGCGCGCGCGCGGCTTAC 793
Qy 220 TyrValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGln 239
Db 794 CCCGTC-----CGAGTGGCGACGACGATCTTTGA-CTGGCAACGAGTTGGCGAG 843

```

Search completed: September 8, 2003, 07:19:43
Job time : 2763 sec

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 07:07:31 ; Search time 343 Seconds
(without alignments)
2762.655 Million cell updates/sec

Title: US-09-889-756A-2
Perfect score: 2019
Sequence: 1 MAFAFKAMRAAAALAAVAL.....APOSQVOTSEAKTASEAE 412

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1537136 seqs, 1149988732 residues
Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-O=/cgnr1/USPRO.spool/US09889756/runat_05092003_124331_23437/app_query.fasta.1.583
-DB=Published Applications NA -OFMT=fastap -SUFFIX=tmpb -MINMATCH=0.1
-LOOPCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62
-TRANS=numan40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HRAPOST=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09889756 @CGN 1 1 244 @runat_05092003_124331_23437
-NCPU=6 -ICPU=3 -NO MMAP -LARGEBUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgnr2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgnr2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgnr2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgnr2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgnr2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgnr2_6/ptodata/2/pubpna/PCUUS_PUBCOMB.seq:
7: /cgnr2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgnr2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgnr2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
10: /cgnr2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgnr2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgnr2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgnr2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgnr2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgnr2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
16: /cgnr2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgnr2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	701.5	34.7	7886	14	US-10-114-170-102

2	435.5	21.6	1395	9	US-09-912-020-121	Sequence 121, App
3	370.5	18.4	1008	11	US-09-975-719-148	Sequence 148, App
4	286	14.2	1830121	14	US-10-329-960-1	Sequence 1, Appl
5	200	9.9	1512	11	US-09-975-719-171	Sequence 171, App
6	175	8.7	1977	10	US-09-738-626-3315	Sequence 3315, App
7	175	8.7	3309400	10	US-09-738-626-1	Sequence 1, Appl
8	173.5	8.6	930	9	US-09-815-242-9752	Sequence 9752, Ap
9	169	8.4	900	9	US-09-815-242-6071	Sequence 6071, Ap
10	162.5	8.0	858	9	US-09-815-242-7815	Sequence 7815, Ap
11	159.5	7.9	1830121	14	US-10-329-960-1	Sequence 1, Appl
12	156.5	7.8	1173	9	US-09-815-242-7049	Sequence 7049, Ap
13	151.5	7.5	1200	12	US-09-769-744A-93	Sequence 93, Appl
14	140.5	7.0	9025608	14	US-10-156-761-93	Sequence 1, Appl
15	138.5	6.9	1068	9	US-09-741-669-242	Sequence 242, App
16	134.5	6.7	3000	10	US-09-855-754-1	Sequence 1, Appl
17	134.5	6.7	3000	14	US-10-227-353-3	Sequence 3, Appl
18	133.5	6.6	4556	9	US-09-784-208-1	Sequence 1, Appl
19	133.5	6.6	4556	14	US-10-315-023-1	Sequence 1, Appl
20	131.5	6.5	2862	14	US-10-156-761-2565	Sequence 2565, Ap
21	131.5	6.5	3000	14	US-10-227-353-5	Sequence 5, Appl
22	131.5	6.5	3116	10	US-09-855-754-3	Sequence 3, Appl
23	130.5	6.5	2398	14	US-10-114-170-260	Sequence 260, App
24	130.5	6.5	25165	14	US-10-114-170-39	Sequence 39, Appl
25	129.5	6.4	1539	10	US-09-938-842A-176	Sequence 176, App
26	128.5	6.4	9025608	14	US-10-156-761-1	Sequence 1, Appl
27	127.5	6.3	1116	11	US-09-988-067B-51	Sequence 15, App
28	125.5	6.2	1794	9	US-09-793-306-145	Sequence 145, App
29	125	6.2	105184	11	US-09-847-513A-1	Sequence 1, Appl
30	122.5	6.1	990	8	US-08-945-038-3	Sequence 3, Appl
31	122.5	6.1	5505	14	US-10-156-761-413	Sequence 413, App
32	122.5	6.1	100000	14	US-10-156-761-15103	Sequence 15103, A
33	121.5	6.0	12145	11	US-09-308-207-19	Sequence 19, Appl
34	121.5	6.0	12145	14	US-10-277-249-1	Sequence 1, Appl
35	121.5	6.0	12145	12	US-10-213-203-1	Sequence 1, Appl
36	120.5	6.0	3543	12	US-10-193-764-60	Sequence 60, Appl
37	120.5	6.0	3568	12	US-10-193-764-58	Sequence 58, Appl
38	119.5	5.9	2088	14	US-10-156-761-1788	Sequence 1788, Ap
39	119.5	5.9	9289	12	US-10-190-435-51	Sequence 51, Appl
40	119.5	5.9	46814	14	US-10-114-170-72	Sequence 72, Appl
41	119	5.9	1116	14	US-10-156-761-5046	Sequence 5046, Ap
42	119	5.9	3695	14	US-10-114-170-18	Sequence 18, Appl
43	118	5.8	2733	10	US-09-855-754-2	Sequence 2, Appl
44	118	5.8	3353	11	US-09-927-827-41	Sequence 41, Appl
45	118	5.8	7755	10	US-09-971-536-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-10-114-170-102/c
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:

```

1  APPLICATION NUMBER: US/10/0114,170
2  FILING DATE: 01-Apr-2002
3  CLASSIFICATION: <Unknown>
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: 09/453,702
6  FILING DATE: 03-DEC-1999
7  APPLICATION NUMBER: 60/110,955
8  FILING DATE: 04-DEC-1998
9  ATTORNEY/AGENT INFORMATION:
10 NAME: Seay, Nicholas J.
11 REGISTRATION NUMBER: 27386
12 REFERENCE/DOCKET NUMBER: 960296, 950117
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (608) 251-5000
15 TELEFAX: (608) 251-9166
16 INFORMATION FOR SEQ ID NO: 102:
17     SEQUENCE CHARACTERISTICS:
18         LENGTH: 7886
19         TYPE: nucleic acid
20         STRANDEDNESS: double
21         TOPOLOGY: linear
22     MOLECULE TYPE: DNA (genomic)
23     SEQUENCE DESCRIPTION: SEQ ID NO: 102:
24 US-10-114-170-102

```

Alignment Scores:	
Pred. No.:	4,84e-65
Score:	701.50
Percent Similarity:	60.10%
Best Local Similarity:	42.78%
Query Match:	34.74%
DB:	14
	Gaps: 7

Oy	7	LysAlaMetAcGValAlaAlaAlaLeuAlaAlaAlaValAlaAlaLeuValLeuSerSerGlySly	26
Db	6823	AAAATAATGAAGACTATATATAGCAACACTCTGTAGAGGCAAGCTGCTTATCTGTGTGT---	6766
Oy	27	LysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGluAlaProAlaProVal	46
Db	6766	-----GATTAATACCCCAAGTAACATTCATCCCGTCAGAAACGAA-----	672529
Oy	47	ValGlyValValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArg	66
Db	6724	GTCCGCGGTGTATACGTCGAATCTCAACCGGCTCTCGTAGTCAGTGAATTAAACCGAGCG	6665E
Oy	67	LeuGlnSerLeuArgThrAlaAspValaArgAlaGlnValGlyGlyIleIleGlnIlySarG	86
Db	6664	ACCAAGTCGCGCTCAGTCCGAGATGACTGCCGAGGTGTGGGGGAATTATCCAGAAACGC	66059
Oy	87	LeuPheGlnGlnGlySerTyrValaArgAlaGlyGlnProLeuTyrGlnIleAspSerSer	106
Db	6604	TTATTATTAAGGAAGGTGATCTGGTCAAGCTGGACGCGCTTCAACAAATTGATCGGCC	65449
Oy	107	ThrTyrGlnAlaAsnLeuGlnUserAlaArgAlaGlnLeuAlaThrAlaGlnAlaPheLeu	126
Db	6544	AGTTATCAGCGCTGCATGGAAATGCAAGCTCGGGCAGCATTTACAAACGACAGGCA---CTG	64886
Oy	127	AlaIlysaAlaAspAlaAspLeuAla--ArgTyrIlySProLeuValAlaAlaGlnAlaVal	145
Db	6487	GTAATAAGCCGATTCGCCAAGAACCGACAGCGTTATGCCCACTCGTGAAGAACGACGCTGT	64282
Oy	146	SerArgGlnGlnTyrAspAlaAlaValThrAlaIlySarGSerAlaGlnAlaGlyAlaIlys	165
Db	6427	TCACAACGAGATCGTATGATGATCTCAAGTCTACCTTGACCAAGATTAAGGCCAGTGTACCG	63686
Oy	166	AlaAlaGlnAlaAlaIleIlySerAlaGlyIleAsnLeuAsnArgSerArgIlePheAla	185
Db	6367	GCGAATAAAGCCGCACTGGAAACTCGCGCATTTATCTTGACTGACGCAACGGTAAACCCCA	63080
Oy	186	ProIleSerGlyPheIleGlyGlnSerIlySerGlyIlyThrLeuLeuAsnAlaGly	205

Db 6307 CCGATTTCCAGGCGGCATTTGGCATTTTCGTCGGTAACCCCTGGACGCATGTCAGCCGCG 6248

Qy 206 AsphTrhValLeuAlaThrIleArgGlnThrAsnProMetIYrValAsnValThGln 225
||||| : : : : :
Db 6247 CAAAGTACAGCGTTAAAGCATTTATTCGGTCTGGATTCAAATGTAATGTCAGCTTCACTTCGC 6188

Qy 226 SerIaSerGIuValMetLysLeuArgArgGlnIleAlaGluGlyLysLeuAlaA 245
||||| : : : : :
Db 6187 TCCAGTGCATTTATTATACGCTTCGCTAAACAG-----TCACCTGGCAACCAAC 6140

Qy 246 AspGIyValIleAlaValGIyIleLysPheAspAspGIyThrValTYrProGluLysGIy 265
: : : : : : : : : : :
Db 6139 AGTGACACCATGAGCGCTCTCACTTATTCGGAAATGGACCAACCTACAGCGAAAAAGG 6080

Qy 266 ArgIeuIeuPheIlaAspProValValAsnGluSerThrGIyGlnIleThrLeuArgAla 285
||||| : : : : :
Db 6079 CGTGTGAAGCTCACCGAAGTCGCTGTAGATGACTACCGGTCGGTACATTCAGGCGC 6020

Qy 286 AlaValProAsnAspGlnAsnIleLeuMetProGIyLeuTYrValArgValLeuMetAsp 305
||||| : : : : :
Db 6019 ATTTTCCCATTCACAACAGCATTTATTACCGGGAATGTTGTTCGCGCTCGTGCAT 5960

Qy 306 GlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGIyAlaLys 325
: : : : : : : : : : :
Db 5959 GAAGGCGGTAGGAAGACGCTATTCTCGCGCCGCAACAGGCGCTCACGCGCGATGCTAAA 5900

Qy 326 Asp-----ThrValMetIleValAsnAlaGlnIlyGlyMetGluProArgGIuValThr 343
: : : : : : : : : : :
Db 5899 GGCATTCGCAACTGCGCTGGTGGTAAATTAAGACATTAAGTAGAGACAGCGCAACCTCGAA 5840

Qy 344 ValAlaGlnGlnGlnGIyThrAsnTrpIleValIleThrSerGIyLeuLysAspGIyAspLys 363
: : : : : : : : : : :
Db 5839 ACGGGAGAAACATATGTGTGATTAATGTGCTGTCTAAACGGCTCGCACAGCGGGACCGCA 5780

Qy 364 ValValValGIuGIyIleSerIleAlaGIyIleThrGlyAlaLysValThrProLys 383
: : : : : : : : : : :
Db 5779 CTGATTGTGAAGGT-----TTCGCCAAAGTCACTTCAGGCGACAGCGTCAAGCTGTT 5726

Qy 384 Glu 384
|||
Db 5725 GAA 5723

```

RESULT 2
US-09-912-020-121
; Sequence 121, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA, 001DVI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ. ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 121
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: E. Coli
US-09-912-020-121
Alignment Scores:

```

Pred. No.: 2,28e-37 Length: 1395
 Score: 435.50 Matches: 124
 Percent Similarity: 47.94% Conservative: 62
 Best Local Similarity: 31.96% Mismatches: 176
 Query Match: 21.57% Indels: 26
 DB: 9 Gaps: 7

US-09-889-756a-2 (1-412) x US-09-912-020-121 (1-1395)

QY 22 LeuSerSerCySgLYLYGLYGLYAspAlaIaIaGlnGlyGly-GlnProAlaGly----- 39
 DB 240 CTCGCCAGTACGAGCCCGAGGCGCAAAACAGCAGCAATTCGCCAGGGGTGCTG 299
 QY 40 -----ArgGlnaIaProAlaIaProValGlyValValThraIHisProGlnTh 56
 DB 300 ACGTGTATGCGTCCGCCCATTAAGCCCGGTTCAGCGCGGAGCCGCCGTAAGAACAGGC 359
 QY 56 rValAlaIaLeuThraValGlnLeuProGlyArgLeuGlnSerLeuArgThraIaAspValAr 76
 DB 360 AGTTCGCGGTACCTCAACCGGCTTGACCATTAACCGCTTAATCCGTTACGCTGCG 419
 QY 76 gAlaGlnValGlyGlyLeuIleGlnLysArgLeuPheGlnGlnGlySerTyValArgAl 96
 DB 420 CAGCCGCTGAGCGCCCACTGATAGCTTACATTTCCAGGAAGCCAGCAGGTCAAGC 479
 QY 96 aGlyGlnProLeuTyrglnIleAspSerSerThraTyrglnIlaAsnLeuGlnSerAlaAr 116
 DB 480 AGCGATTTACTGGCAGAAATTAAGCCCGACGTCATTAATTCATTAAGCAGCAAGCCCA 539
 QY 116 gAlaGlnLeuAlaThraIaGlnAlaThraLeuAlaLysAlaAspLeuAlaArgTy 136
 DB 540 GGGCCAACTGGCAAAAGATTAAGCCAGCTTCCCAAGCCCGCTGACCTGCGCTTA 599
 QY 136 rLYeProLeuValAlaIaIaGlnAlaValaIaSerArgGlnGlnTyraAspAlaIaValThraI 156
 DB 600 TCACCACTGGCAAAACCAATCTCGTTTCCCGAGAGCTGGATGCGCAACAGCGGCT 659
 QY 156 aLYsArgSerAlaGlnAlaIaGlyValLysAlaIaIaAlaIaIaLysSerAlaGlyI 176
 DB 660 GGTCAATGAAACCGAAGCAGCATTAAGCTGATGAAGCAAGCTTCCAGCGGAGCT 719
 QY 176 eAsnLeuAsnArgSerArgIleThraIaProIleSerGlyPheIleGlyGlnSerLysVa 196
 DB 720 GCAATCGATCGAGCGCGGATTAACCGCAGCTGATGCTGCTGCTCAAGCAGGT 779
 QY 196 lSerGlnGlyThraLeuAsnAlaGlyAspThraThraValLeuAlaThraIleArgGlnTh 216
 DB 780 TGATGTTGGTACCAAAATCTCCAGTGTGATACCAAGCGGATCGTGTGATCACCGAGC 839
 QY 216 rAsnProMetTyraValaIaIaThraIaSerAlaSerGlnValaIaMetLysLeuArgAl 236
 DB 840 GCATCTATGATTAAGTCTTACCTTCGCCGAAGAGATATGCTTCCGTAAGTCAGGC 899
 QY 236 nIleAlaGlnGlyLysLeuLeuAlaIaAspGlyValIleAlaValaGlyIle--LysPh 255
 DB 900 GCAGAAACCGGAAACCGCTGTGTGTAAGAGCTGGGATCGCACAACCGAAGAAAT 959
 QY 255 eAspAspGlyThraValTyProGlnLysGlyArgLeuLeuPheAlaAspProValVala 275
 DB 960 AAGTGAAGGAGCG-----CTGTAAAGTCTAATTAACCAAAATCGA 998
 QY 275 nGlnSerThraGlyGlnIleThraLeuArgAlaIaValaIaProAsnAspGlnaIleLeu 295
 DB 999 TGGCACATCCGATGATTAAGTGAAGAGCAGCTTTAAATATAGATATGCTGCTT 1058
 QY 295 tProGlyLeuTyraIaIa-----ArgValaIaLeuMetAspGlnValaIaValaAspAlaPh 313
 DB 1059 TCCCAATCAAGTTGTAACCGCGCATGTTAGTGCAC-----ACCGAACAAGCGCGT 1112
 QY 313 eValaIaIaProGlnAlaIaIaThraArgGlyAlaLys--AspThraValaIaIaVala 332
 DB 1113 AGTATCCCAACAGCGCGCTCGCAATATGAGCAATGAAGGCCATTTGTTGCGGTGAA 1172

QY 332 nAlaGlnGlyGlyMetLeuProArgGlnValaIaThraValaIaGlnGlnGlnThraAsnTr 352
 DB 1173 TAGCAAAACAGAGTACCAAAACATCTGTGAGCGCGGATTCAGACATCGAAAGT 1232
 QY 352 pIleValThraSerGlyLeuLysAspGlyAspTyraValaIaGlnGlyIle----- 369
 DB 1233 GGTGATCCGTGACAGTATTTCTGCGGGCGATCGCGGTGTACAGACGCGCATGATCGCT 1292
 QY 370 -----SerIleAlaGlyIleThraGlyAlaLysLysValThraProLysGlnTr 385
 DB 1293 GACCGAAGGCGCAAAAGTGAAGTGTGGAAGCCAGAGCCCATCTCCGAAGAGAA 1352
 QY 385 pAlaSerSerGlnaIaIa 392
 DB 1353 AGCCACAGCGCGAATATACCG 1374

RESULT 3
 US-09-975-719-148
 ; Sequence 148, Application US/09975719
 ; Publication No. US2003022349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ausubel, Frederick M.
 ; APPLICANT: Rahme, Laurence G.
 ; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
 ; FILE REFERENCE: 00786/361003
 ; CURRENT APPLICATION NUMBER: US/09/975,719
 ; CURRENT FILING DATE: 2001-10-10
 ; PRIOR APPLICATION NUMBER: US 09/199,637
 ; PRIOR FILING DATE: 1998-11-25
 ; PRIOR APPLICATION NUMBER: US 60/066,517
 ; PRIOR FILING DATE: 1997-11-25
 ; NUMBER OF SEQ. ID NOS: 437
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ. ID NO 148
 ; LENGTH: 1008
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-975-719-148

Alignment Scores:
 Pred. No.: 1.51e-30 Length: 1008
 Score: 370.50 Matches: 80
 Percent Similarity: 58.29% Conservative: 43
 Best Local Similarity: 37.91% Mismatches: 73
 Query Match: 18.35% Indels: 15
 DB: 11 Gaps: 5

US-09-889-756a-2 (1-412) x US-09-975-719-148 (1-1008)

QY 205 GlyAspThraThraValaIaIaThraIleArgGlnThraAsnProMetTyraValaIaThra 224
 DB 1 GGCAGGCAAAACCGATGGCCAGCGCAACACCTGACCCGATCTACGTCGACGTAC 60
 QY 225 GlnSerAlaSerGlnValaIaIaMetLysLysArgGlnIleAlaGlnGlyLysLeu--Leu 243
 DB 61 CAGCCGTCAACCGCCCTGTTCGATGCGCCGCGAACTGCGCAGCGCGCATTTGAGCGC 120
 QY 244 AlaAlaAspGlyValIleAlaValaGlyIleLysPheAspAspGlyThraValTyProGln 263
 DB 121 GCCGCGCAACAGCTCGAAGGTCCTCGAAGCTGGAAGAGCGGTGACCAATACCGCTG 180
 QY 264 LysGlyArgLeuLeuPheAlaAspProValaIaAsnGlnSerThraGlyGlnIleThra 283
 DB 181 GAAGGCGCTCGAATTCCTCGAGGTTTCGTCGAGGAAGGACCGGCTCGTCAACATC 240
 QY 284 ArgAlaAlaValaIaProAsnAspGlnaIleLeuMetProGlyLeuTyraValaIaArgVala 303
 DB 241 CCGCGCGGTGTCGCCAACCCGAAACGAGCTGCTCCCGCATGCTTCACGCGCAG 300
 QY 304 MetAspGlnValaIaValaAspAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 323
 DB 301 TTGACAGGAAGCGTCAAGCAGAAAGCCATCTGCTCCGACAGGAGCGTGAACCGCGAGC 360

```

Oy      32  A l y s e s p -----ThrValMetCileValAlaAnalagInglYlmeGluProArgGlu 341
Db      361  C T A A G G G C C A G G C T A C C G C G C G T G G T G A A C G G C A A C A A G C T G A C C T G C G G T G 420

Oy      342  V a l T h r V a l A l a G l n G l n G l n G l Y t h r A n T P l l e v a l T m S e r G l y L e u l y e a s p G l y 361
Db      421  A T C A A G G C C G A C C G G T G A T C G G C G C A A G T G G C T G T C A C G A A G G C C T A A C C C G C G C 480

Oy      362  A s p l y e v a l V a l A l u G l u G l y l e s e r l l e a l a -----G l y l l e t h r G l y A l a l y s l y s 379
Db      481  G A C A A G T C A T T T A C C A A G C C C T G C A G T T C G T C A C C G C G G T G T G A C ---G T G A A A A C C 533

Oy      380  V a l T h r P r o l y s G l u T r p A l a s e r S e r G l u a s e n G l n a l a l a l a P r o G l n S e r G l y V a l 399
Db      538  G T G C C G G C G A A G A T G C G C G T C G G C G A A G G C C G A C G C G C T C C G -----585

Oy      400  G l n T h r A l a s e r G l u a l A l y s e r T h r A l a s e r G l u 410
Db      586  -----G C G A A A A C C A C A C A C A A G 603

```

```

RESULT 4
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US200300992772A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:

```

NAME/KEY: misc.feature	
LOCATION: (40808)..(40810)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (44416)..(44416)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (44905)..(44905)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (44975)..(44975)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (44593)..(44593)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (45732)..(45732)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (47036)..(47036)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (51334)..(51334)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (51602)..(51602)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (51786)..(51786)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (51805)..(51805)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (55369)..(55369)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (65309)..(65309)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (65313)..(65313)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (80024)..(80024)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (100091)..(100091)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (102596)..(102596)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: (105121)..(105121)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc.feature	

```
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 5.98e-17 Length: 1830121
Score: 286.00 Matches: 105
Percent Similarity: 41.01% Conservative: 66
Best Local Similarity: 25.18% Mismatches: 148
Query Match: 14.17% Indels: 98
DB: 14 Gaps: 14

US-09-889-756a-2 (1-412) x US-10-329-960-1 (1-1830121)

QY 6 PheylsalmeatrgalaalaleuAlaAlaValAlaLeuValLeuSerSerCys 25
Db 946822 TTTAATATGATTAATAAGCGTAATGATTAAGCCGAGCTTCA----- 946863

QY 26 GlyLysGlyGlyAspAlaAlaGlnGlyGlyProAlaGlyArgGlnAlaProAlaPro 45
Db 946864 -----GGAATGCCA-----GAAATCTTCAAGCCCA 946887

QY 46 ValValGlyValValThrValHisProGlnThrValAlaAlaLeuThrValGluLeuProGly 65
Db 946888 GTAACC---GCACCTGAAGTTCAACCGCGTAATGACGACGACGACGATTATTAACAACAACAGT 946944

QY 66 ArgLeuGluSerLeuArgThrAlaAspValAlaArgAlaGlnValGlyGlyIleIleGlnLys 85
Db 946945 CTGTGGTCCTCAATCAAGCGCAATGCTCAGTACACAAATGCGGCGGCTTCAAA 947004

QY 86 ArgLeuPheGlnGluGlySerTyrrValAlaArgAlaGlyGlnProLeuTyrrGlnIleAspSer 105
Db 947005 GTACTTGTTCAAAATGACCAAAATGTCAAAAGAGTGAGTGCTTGGAGCTTGAATGT 947064

QY 106 SerThrTyrrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
Db 947065 TCTGTGAACGACCTATCTACAGCTGCTCAGGCAATTAACAGACCTCGCAAACT 947124

QY 126 LeuAlaLysAlaAspAlaAspLeuAlaArgTyrrLysProLeuValAlaAlaGluAlaVal 145
Db 947125 -----TACCAACGTTATGTGGTTATTAATACCAATGCTGTA 947163

QY 146 SerArgGlnGluTyrrAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLys 165
Db 947164 TCACGTCAAAATGATGATTAACGAAAGCGCTTAATGATGCTCACTAGCTAGTATGAA 947223

QY 166 AlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
Db 947224 TCTCTAAAGCGACCAATTGA-----CGTGGTAAATTTGTTGCG 947262

QY 186 ProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGly 205
Db 947263 CCATTGATGCGCAACAGATATTGTGAATCAATGTTGACAAATATGTAATGTTGGA 947322

QY 206 AspThrThrValLeuAlaThrIleArgGlnThrAspProMetCysTyrrValAsnValThrGln 225
Db 947323 -----ACGAAATTTGCTGCTGAGAAAGATACCTCAATGAAGATGATTTGCTCTT 947376

QY 226 SerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGlyLysLeuLeuAlaAla 245
Db 947377 TCACAAATGATTTAATAAATTACAT-----ATCGGTACAGCGCTTACACGCG 947424

QY 246 AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrrProGluLysGly 265
Db 947425 ACAACAGATGCTGCTGCGCGGAACATTT-----TCAGCT 947460

QY 266 ArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuAsnAla 285
Db 947461 CGAATACGCGCATTTGACCTGCAATTAATCAACAGGTTTACTGATGTTCAAGGCT 947520

QY 286 AlaVal---ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrrValArgValLeuMet 304
Db 947520
```

Db 947521 ACTTTGATCCCTGAGATGGGCAATTAATGCTTCAGGTATGCTCTCGCTTACGCANT 947580
Qy 305 AspglnValAlaValAspAsnAlaPheValValProglnglnAlaValThr----- 321
Db 947581 GCACTTCCAACTGAAACAAATCAAGTGTGTCACAGTAGCTATTAGCTACAAATATG 947640
Qy 321 ----- 321
Db 947641 TATGGCGAATTTGCTATTACTTGAACCATTAATCTGAAGAAAGAAAGAAATATGCA 947700
Qy 322 -----ArgGlyAlaValAspThrValMetIleValAsn 332
Db 947701 GGTAATGAAATAATGATGCTCTCTATGTCGAAACAGATCCGATTTTACTTAAGAT 947760
Qy 333 AlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyr 352
Db 947761 CGTCAAGGT-----GTTATGCTCAATTACAGGAAATGAA--- 947796
Qy 353 IleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIleAla 372
Db 947797 -----GTTAAAGTGGAGATATAAATTATACAGCGCGTCACGAA----- 947835
Qy 373 GlyIleThrGlyAlaLysLysValThrProLysGluTTPalAspSerGlu 389
Db 947836 GGATTTGGTAAATGGAAGTCTTGTTG-----GATGCAATTAAGAAAGAC 947877

RESULT 5
US-09-975-719-171/c
; Sequence 171, Application US/09975719
; Publication No. US20030022349a1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 171
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-171

Alignment Scores:
Pred. No.: 6,65e-12 Length: 1512
Score: 200.00 Matches: 47
Percent Similarity: 65.93% Conservative: 13
Best Local Similarity: 51.65% Mismatches: 25
Query Match: 9.91% Indels: 6
DB: 11 Gaps: 3

US-09-889-756a-2 (1-412) x US-09-975-719-171 (1-1512)

Qy 14 LeuAlaAlaValAlaValAlaLeuValLeuSerCysGlyGlyGlyAspAlaAlaGln 33
Db 255 CTGGGGCGCCCTGGCTGCTTCTCTGCTGGCTC-----GAAGAACGACGCGAC 205
Qy 34 GlyIleGlnProAlaGlyArgGluAlaProAlaProAlaValGlyValValThrValHis 53
Db 204 ACCGGAGAGCTGCG-----GAGGCCCGCGCGAG-----GTGCGCGTATGCTGCGCAG 154
Qy 54 ProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAla 73
Db 153 CCGGGCCGATCGGATCCAGCGAGCTGCGCGAGCCCTGGAGCGCTACCGCCAGGCC 94
Qy 74 AspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyr 93

Db 93 GAGGCGCGCGCGCGCGCTGCGCGCATGCTACCCCGTGCCTGTACGAGAGGCGCAGCA 34
Qy 94 ValArgAlaGlyGlnProLeuTyrGlnIleAsp 104
Db 33 GTCCGCGCGCGCACCGTGTCTTCCAAATGAC 1

RESULT 6
US-09-738-626-3315
; Sequence 3315, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKADA, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3315
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3315

Alignment Scores:
Pred. No.: 4.8e-09 Length: 1977
Score: 175.00 Matches: 102
Percent Similarity: 39.70% Conservative: 81
Best Local Similarity: 22.13% Mismatches: 194
Query Match: 8.67% Indels: 84
DB: 10 Gaps: 16

US-09-889-756a-2 (1-412) x US-09-738-626-3315 (1-1977)

Qy 5 AlaPheLysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSer 24
Db 667 GCTCTCAAGGACGCTGCTGAGCAAGCAGATGCTCTGAATTTGACAGCATCGCGCGGAC 726
Qy 25 CysGlyLysGlyGlyAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAla 44
Db 727 TTGGAGCTGCTCAACAC-----GATCCAGCAACCTCAAC 762
Qy 45 ProValValGlyValValThrValHisProGlnThrValAlaLeuThrValGlnLeu--- 63
Db 763 GATGCTCATTTGCTGCTGCGATGAGAGAAATCTTTGCTTCGCGAATCCGAATCGAATGCT 822
Qy 64 -----ProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 823 CAAGCCCGCGCTGAGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
Qy 80 GlyIleIleIleGlnLysArgLeuPheGlnGlnGlnGlySerTyrValArgAlaGlyGlnPro 99
Db 877 GAGAGCTGAGACATCATTAATGCTTCCAAACCTCCACAGTGGCTTACCAAGATCAGACT 936
Qy 100 LeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
Db 937 TACTTGAG-----TCTTACACCGCTTTGGAGGAGAGAGAGAGAGAGAGAGAGAGAG 984


```

Qy 120 AlaThrAlaGlnAla-----ThrLeuAlaLysAla 129
Db 985 TCCACCACTGAAGCTCTGAAATAGCCGAGGATCTACATTGATTCACTGGAAAAGTT 1044
Qy 130 AspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGln 149
Db 1045 GACTCAGAACTAGCCCGCCGACGCGCCGCTT---GCCGAGGCCACTCGACCAACAA 1101
Qy 150 TyrAspAlaAlaVal-----ThrAlaLysArgSerAlaGluAlaGlyValLysAla 166
Db 1102 ---GACGACGACCTTGGCTCGAAGCCGCGACGCTTCCACCAACCAATGGAAGCC 1158
Qy 167 AlaGlnAlaAlaLysSerAla----- 174
Db 1159 CAATCAAGCCGATCGATGACGCTTAGTTGGATCAGTAAATGAAGCCGCCACC 1218
Qy 175 -----GlyTLeuLysLysLysSerArgGlnThrAlaProLysSer 188
Db 1219 AGATCCACGTCCTCAGCTCGGATGATCAACCAACACCCGCTCGCTCCCATCTCA 1278
Qy 189 GlyPheLe-----GlyLysSerLysValSerGlnGlyThrLeuLys 203
Db 1279 GGCATTGTTTATCCGTCGAGCAGCCAGGTCAACGACGCGCGCATCTTTAGT 1338
Qy 204 AlaGlyAspThrThr-----ValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
Db 1339 GTTCTGATGATTCGGAAGTCAAGATCAAGCCGATGTAAGAAAGAGGAGATGACAA 1398
Qy 221 ValAsnVal-----ThrGlnSerAlaSerGluValMetLysLeuArg 234
Db 1399 GTCAACCATCGGATCCCGCTCACCTTCACTCCCATCGGACCAAAAGATTCGCC 1458
Qy 235 ArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyLys 254
Db 1459 GCGCGAGATCCAAAGTCTCCCATTCGACGCT----- 1491
Qy 255 PheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAlaAspProValVal 274
Db 1492 -----GCCGCACTGCCCGCCAGCTACAGCTAAGAGCTGCCCGACGCCCAAC 1542
Qy 275 AsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeu 294
Db 1543 ACCAACAACGTCACCTTCCCATGAAATTTCCGTACCGGCGACCGCAAGGCTTC 1602
Qy 295 MetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheVal 314
Db 1603 AACCTCGCGGATCCGCTCGAGTACGATCGCTCATGAATCGACACCACTGAC 1662
Qy 315 ValProGlnGlnAlaValThrArgGly-----AlaLysAspThrValMetIleValAsn 332
Db 1663 GTTCTCTTGAAGCTGTGTACAAATGATGACGCGCAAGACGCTGTTTGAATCATAGC 1722
Qy 333 AlaGlnIleGlyMetGluProArgGluValThrValAlaGlnGlnGlnIleThrAsnTyr 352
Db 1723 GACGACAAACAAGTAAAGTAAAGTAAACACGCTGATCCGATGCTTGAATATC 1782
Qy 353 IleValThr---SerGlyLeuLysAspGlyAspLysValValGluGlyLysIleSerIle 371
Db 1783 GCAGTACGCGGTGGAATTTCAAGAGACGCTGACGCTCCCGCTGGAATATC 1842
Qy 372 AlaGlyIleThrGlyAlaLysValThrProLysGlnThrAlaSerSerGluAsnGln 391
Db 1843 CGGGGCTCTACGGA---ACTGTGAATCTTACGACGACGATGCGTGAAGCAG 1893
Qy 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGlnAlaLysThrAlaSerGluAla 411
Db 1894 GCGCGGCTCTTTAGTCTCGCGGCTTTGACCTGACGCGCTCCGCTTTCGCC 1953
Qy 412 Gln 412
Db 1954 AAG 1956

```

```

RESULT 7
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

```

Alignment Scores:
Pred. No.: 0.000129 Length: 3309400
Score: 175.00 Matches: 102
Percent Similarity: 39.70% Conservative: 81
Best Local Similarity: 22.13% Mismatches: 194
Query Match: 8.67% Indels: 84
DB: Gaps: 16

```

```

US-09-889-756a-2 (1-412) x US-09-738-626-1 (1-3309400)
Qy 5 AlaPheLysLeuMetArgAlaAlaAlaAlaAlaValAlaLeuValLeuSerSer 24
Db 3196520 GCTCTCAAGGACCTGTGAGCAAGCATGCTGCAATTAAGCATGCCGCGAC 3196461
Qy 25 CysGlyLysGlyAlaAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAla 44
Db 3196460 TTGGACCTGCTCAACAC-----GATGGACGACACCTCAAC 3196425
Qy 45 ProValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeu--- 63
Db 3196424 GATGTCATTGCTCTGCTCATGAGAGAAATCTTGGCTTCTCGGAATCCGAATGCT 3196365
Qy 64 -----ProGlyArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 79
Db 3196364 CAAGCCGCGCTGACGATGACGAGGCGATGCTGACGCTGAA-----GCAAGGTT 3196311
Qy 80 GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 99
Db 3196310 GCAAGGCTGGAACATCATCTGCTTCCAAACCTCAGCTGCGTGAAGTCAACT 3196251
Qy 100 LeuTyrGlnIleLysSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
Db 3196250 TACTTCAG-----TCTTACACCGCTTTCGAGGAAGCTGAGCGACGCTGCA 3196203
Qy 120 AlaThrAlaGlnAla-----ThrLeuAlaLysAla 129
Db 3196202 TCCACCACTGAAGCTTGAATAGCCGAGGATCTACATTGATTCACTCGGAAAAGTT 3196143
Qy 130 AspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGln 149

```

```

Db 3196142 GACTCAGAACTAGCGCCGCGACGCGCGCT---GCCGAAGCCCACTCAGCAACA 3196086
Qy 150 TyrAspAlaAlaVal-----ThrAlaYarSerAlaGluAlaGlyValIleSera 166
Db 3196085 ---GACGAGACACTTGGCTCGAAGACCGCGACAGCTTTCCACCACCAATTTGAAACC 3196029
Qy 167 AlaGlnAlaAlaIleYsSerAla----- 174
Db 3196028 CAATCAAGCGCCATCGATCGAGCTTTAGGTTGGCATCATGATTAATGAAGCCGCCACC 3195969
Qy 175 -----GlyIleAsnLeuSeraIleThrAlaProIleSer 188
Db 3195968 AGATCCAGTCCCGACGTCGGATGATTCACACACACCGCTCGCTCCCATCTACTCA 3195909
Qy 189 GlyPheIle-----GlyIleSeraValSerGluIleThrLeuAsn 203
Db 3195908 GGCATGTTTATTCATCCGTCGACGACGCCAAGTCAACACGACGCGCGCATGTTGAGT 3195849
Qy 204 AlaGlyAspThrThr-----ValIleuAlaThrIleArgIleThrAspProMetTyr 220
Db 3195848 GTTGCTGATGATTCGGAAGTCAAGATCACCGGATGATTAAGAAAGCGAGATCGCAAC 3195789
Qy 221 ValAsnVal-----ThrIleSeraIleSerGluValMetIleSeraArg 234
Db 3195788 GTTCAACATCGGATCCCGGTCACCTTCACTCCCATCGGAAACCAAGATTCGCC 3195729
Qy 235 ArgGlnIleAlaGlnGlyIleValLeuAlaIleAspGlyValIleAlaValIleVal 254
Db 3195728 GCGCGAGATTCCAAGTCTCCCATTCGACCT----- 3195696
Qy 255 PheAspAspGlyThrValTyrProGluIleSeraIleValIleAspProValVal 274
Db 3195695 -----GCCGCGAGTCCCGGTCAGTCACTACGTCAGAGAGTCCGCGCGCAACAAC 3195645
Qy 275 AsnGluSerThrIleGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleu 294
Db 3195644 ACCAACAAGTCACTGACCTTCCCATCGAAATTTCCGTCACCGCGAGCGGAAAGGCTTC 3195585
Qy 295 MetProGlyLeuTyrValArgValIleuMetAspGlnValAlaValAspAsnAlaPheVal 314
Db 3195584 AACCTCGCGGATCCGCTCGAGTACGATCTCCATGAAATCGCACCACTCATCTAC 3195525
Qy 315 ValProGlnIleAlaValThrArgGly-----AlaYsAspThrValMetIleValAsn 332
Db 3195524 GTTCTCTTGAAGGCTGTCAAAAATGATGACGCGCAAGCGCTTTGATCATCAGC 3195465
Qy 333 AlaGlnGlyIleMetGluProArgGluValThrValAlaGlnGlnGlnIleThrAsnTrp 352
Db 3195464 GACGACAAACAAAGTAGAAGTAGAAGTAAACAGCTGAATCGATGACTTGTGATTC 3195405
Qy 353 IleValThr-----SerGlyLeuIleuAspGlyValValIleGluIleSeraIle 371
Db 3195404 GCACTACGCGGTGCTGCAATTTTCAGACAGCGTCAGTGCACCCAGCTCGAAACTAC 3195345
Qy 372 AlaGlyIleThrGlnIleValIleValThrProIleSeraIleAspSerGluAsnGln 391
Db 3195344 CGGGCGCTCATCGGAGAA-----ACTGGAATCTTACGACGATACGCTGAGAGAG 3195294
Qy 392 AlaAlaAlaProGlnIleSerGlyValIleThrAlaSerGluAlaIleValIleSeraIle 411
Db 3195293 GCGCGCGCTCTTTAGTCTGCGGCGCTTTTGAACCTGACGCGCGCTTCTGCC 3195234
Qy 412 Glu 412
Db 3195233 AAG 3195231

```

```

APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9752
LENGTH: 930.
TYPE: DNA
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(930)
NAME/KEY: misc_feature
LOCATION: (1)..(930)
OTHER INFORMATION: n = A,T,C or G
US-09-815-242-9752
Alignment Scores:
Pred. No.: 2,47e-09 Length: 930
Score: 173.50 Matches: 70
Percent Similarity: 42.11% Conservative: 50
Best Local Similarity: 24.56% Mismatches: 117
Query Match: 8.59% Indels: 48
Gaps: 8
US-09-889-756a-2 (1-412) x US-09-815-242-9752 (1-930)
Qy 48 GlyValValThrValIleProGlnThrValAlaIleuThrValGluLeu-ProGlyArgLe 67
Db 129 GCGCGCGTGGGTATGGAATTAATCTATGCACTGACCGTGGACCGCGCATGTAAT 188
Qy 67 uGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnIleSeraIle 87
Db 189 CCGCGCG---GAACAGCTGACGTAACGCCACGAGTTTCTGGCAGCATACCCAACTCAA 245
Qy 87 uPheGlnGluGlySerTyrValArgAlaGlnIleProLeuTyrGlnIleAspSerSerTh 107
Db 246 TATCAAAAGATATATGATGTTGCAACGCTGGGAGCGTTCTTGTCTATGATAAAAGCC 305
Qy 107 rTyrGluAlaAsnLeuGluSerAlaArgAlaGlnIleuAlaThrAlaGlnAlaThrLeuAl 127
Db 306 CTTCATATTTGCAAGACTTAACGGGACGACAGCTTGGCCAAAGCGCATGCTCGGC 365
Qy 127 alyVala-----AspAlaAspLeuAlaArgTyrIleProLeuValAlaAlaGluAlaVa 145
Db 366 AAAAGCGAATTAATGAACGATGACCTCGACAT-----TTATCGCGTAACTACTAT 416
Qy 145 lSerArgGlnGluTyrSerAspAlaAlaValThrAlaYarSerAlaGluAlaGlyValIly 165
Db 417 CTGTGCTGAAGATCTGATGACCGCATATGTAATGTTAAAGCATGACGACCGCATGA 476

```

```

QY 165 SALAAlaGlnAlaAlaIlelySerAlaGlyIleAsnLeuAsnArgSerAlaIleThrAl 185
Db 477 TGTCGCACTGGCAACGTTAAACAGCGCAGTGGCAGCTTAGCCAGACGGAATGAAAC 536
QY 185 AProIleSerGlyPheIleGlyGlnSerIleValSerIleGlyThrIleLeuAsnAlaG 205
Db 537 GCCGGATCCCGCTGGGAGCACCACCTTCCACCGGATCCGATGATTATGCTACGACCGG 596
QY 205 YASPThrValIleuAlaThrIleArgGlnThrAsnProMetIleValAsnValThrG 225
Db 597 T-----AAGCCACTTTCCGCGCTGATGATAG 623
QY 225 nSerAlaSerGluValMet-----LysLeuArgArg----- 235
Db 624 CCATTCATTATTAATGGATATTTGCAAGAAACCAATACGACCATATCCGCAAG 683
QY 236 -----GlnIleAlaGlnGlyIleLeuLeuAlaAlaAspGlyValIleAl 250
Db 684 TGACCGCTGATTAATACACGTATAGTGCATATGTTAAGTACAGGCTCAGTTGGCAG 743
QY 250 aValGlyIleLysPheAspAspGlyThrValIleThrLeuArgAlaAlaValProAsnAs 270
Db 744 TATCGGGGGCGCTTTTACGACCAAGCGTGAAGTACCTCCGTTTATG----- 795
QY 270 aAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAs 290
Db 796 -----CCGATTCATCAACCAACGACGCTGGGTAGCTGGCGCAACGCTTACGG----- 846
QY 290 pGlnAsnIleLeuMetProGlyLeuIleValArgValLeuMetAspGlnValAlaValAs 310
Db 847 -----GlnGATGCAATTTTACGCTTTGCGCAGGA 878
QY 310 pAsnAlaPheVal 314
Db 879 TATCAGCGCTGCTC 891

```

RESULT 9

```

US-09-815-242-6071
Sequence 6071, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6071
LENGTH: 900
TYPE: DNA

```

```

; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (900)
US-09-815-242-6071
Alignment Scores:
Pred. No.: 7,23e-09 Length: 900
Score: 169.00 Matches: 71
Percent Similarity: 40.86% Conservative: 43
Best Local Similarity: 25.45% Mismatches: 105
Query Match: 8.37% Indels: 60
DB: 9 Gaps: 9

```

US-09-889-756a-2 (1-412) x US-09-815-242-6071 (1-900)

```

QY 52 ValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArg 71
Db 160 ATACGGCGCAGAACAGATTCTTATACCCACAGGTGCGGAGCGATATTGTGAGCTGAAT 219
QY 72 ThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlnGly 91
Db 220 ATAAAGAC-----AAC 231
QY 92 SerIleValArgAlaGlyGlnProLeuIleAspSerSerThrIleArgIleAlaAsn 111
Db 232 CAGCTGGTTAACCGGGGATCTTCTGCTCAGATCGATGAATAAAGCCCTTACAGATGCC 291
QY 112 LeuGluSerIleArgAlaGlnLeuAlaThrAlaGlnAlaThrIleuAlaValAlaAspAla 131
Db 292 GAGCTGACGACAGAGCTCAGCTGGCGAAGCCCAATCTGACCTCGCAAAAGCCAAATAC 351
QY 132 AspLeuAlaArgIleThrLysProLeuValAlaAlaGlnValAlaSerArgGlnIleIleAsp 151
Db 352 GAAGCGAATGCTGCTGCTATCTC---TCACAAAATTTATTTCTGCCGAAGACTCGAT 408
QY 152 AlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIle 171
Db 409 ACCGTTAACCTCAATGTTAAACGATGACAGCGCAGGCTGATGATCCGACAGGAGCGTG 468
QY 172 LysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIle 191
Db 469 AAACAGGCAACATGCAATGCGGCAACAGCAAAATTCGGCCCGGTGATGATGGGTG 528
QY 192 GlyGlnSerIleValSerGlyThrIleLeuAsnAlaGlyAspThrThrValLeuAla 211
Db 529 ACTAACCTTACTACCGCATCGCGATTAACGTTACCGG----- 570
QY 212 ThrIleArgGlnThrAsnProMetIleValAsnValThrGlnSerAlaSerGluValMet 231
Db 571 -----AAACCGCTTTGCCCTTGTGACAGCACCATCGTTTACGTCATT 615
QY 232 -----LysLeuArgArgGlnIleAlaGlnGlyLys----- 241
Db 616 GGTATTATTGAAGAAACCAAGTTG---CCCATATCCGGAAGTGCACACAGCAAAATT 672
QY 242 LeuLeuAlaAlaAspGly-----ValIleAlaValGlyIleLysPhe 255
Db 673 ACCCTTATAGTATTAACAAAACGTTACAGGCTCAGCTTCCAGTATCGGTCCGGCGATT 732
QY 256 AspAspGlyThrValIleProGluLysGlyArgLeuLeuPheAlaAspProValValAsn 275
Db 733 TATGATCAGAGCGTTTAAAGTACGTAACCTTATT-----CCGATGTTAAA 780
QY 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMet 295
Db 781 CCTAACGTCCTCGGATTCGACTCGCCCAACCGTTCC----- 819
QY 296 ProGlyLeuIleValArgValLeuMetAspGlnValAlaValAspAsnAlaPheVal 314
Db 820 -----GTCCGCTTGGCGCTGGATTAAGTCCAGCGATGTCACTGTTGGTG 864

```

RESULT 10


```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c

```

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10091)..(10091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (12167)..(12167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```

Db 954023 ATGGCAGTGTATCTAATGAACAAATGTGGCTAGAGCTAACTTCAAGAAGAACCCCAATTA 953

```

RESULT 12
US-09-815-242-7049
Sequence 7049, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haasebeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7049
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1173)
US-09-815-242-7049

Alignment Scores:
Pred. No.: 2,32e-07 Length: 1173
Score: 156.50 Matches: 71
Percent Similarity: 37.20% Conservative: 51
Best Local Similarity: 21.65% Mismatches: 115
Query Match: 7.75% Indels: 91
DB: 9 Gaps: 10

US-09-889-756a-2 (1-412) x US-09-815-242-7049 (1-1173)

Qy 75 ValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlySerTyrVal 94
Db 190 GTATCATCACAAGTTGACAGTAATGACGAAGATTATGACAGCAATATGATTAAGTC 249
Qy 95 ArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSer 114
Db 250 CATCAGGTGATTAATCTTCAAGTGAATGGATGATCAATGCGAATCTTGAATTCGACAG 309
Qy 115 AlaArgAlaGlnLysAlaThrAla 122
Db 310 GCCAAAGCAATCTGCCAATGCTGTACGTACGAGCAACCTGGGTTTCACTGTGACG 369
Qy 123 -----GlnAlaThrLeuAlaLysAlaAspAlaAspLeu 133
Db 370 CAATTGCAATCAGCTGTCAGCAGCAATGAATTTCTTAGCTCAAGCTCAAGCAATTTG 429
Qy 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnLysTyrAspAlaAla 153
Db 430 GCGGCGCAGAGTTCAACTGTAAGAAATGGGTGCGATTGTAAGATCTTTCACACGCA 489
Qy 154 ValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGln 168
Db 490 AAGAGAGCTGTCGAGCTTGCACAAAGCAATTTAAATGCTCTAAATCAATTGGCTGCT 549
Qy 168 ----- 168
Db 550 AATCAGCTTATTCGTAACGTTCCATTACGTGACGCAACCGCAATATACGAATCGATA 609
Qy 169 AlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSer 188
Db 610 AACTCGCTTAAACAGGCTTGTTAAATTTGCAACGTACCAAAATTTAGATCCCAATTGAT 669
Qy 189 GlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThr 208
Db 670 GGGCATATGCGCGCTCTTAATGTTCAAGTTGCGGACCGCTTTCAGTGGGTGGCGCTTA 729
Qy 209 ValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSer 228
Db 730 ATGCGCATGCTATTAATGAACAAATGTGGCTAGACGTACCTTCAAGAAACCAATTA 789
Qy 229 GluValMetLysLeuArgArgGlnIle-----AlaGlnGlyLysLeuLeu 243
Db 790 ACGAATATGCGCATTTGTCAGCCAGTAAATCAATTTGATTATACGGTAAAAATAAA 849
Qy 244 AlaAlaAspGlyValIleAlaValAlaGlyIleLysPheAspSerGlyThrValTyrProGlu 263
Db 850 GAATTTATGCGCGATTAAT--GGTATTGAATGGGAGCTGGCAATGCAATTT----- 900
Qy 264 LysGlyArgLeuPheAlaAspProValValAsnGlnSerThrArgLysGlnIleThrLeu 283
Db 901 -----TCTCTTTTG-----CCTTCACAAATATCTACGGGTAACTGGATTAAGTG 945
Qy 284 ArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeu 303

```

```

Db 946 GTGCAACCGTACCT-----GTGCAATTAA 972
Qy 304 MetAspGlnValAlaValAspAsnAlaPheValValProGlnGln----- 318
Db 973 TTACAT-----CCACAACAAATTACGAACCAACGCCA 1002
Qy 319 --AlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyMet 337
Db 1003 TTGAGAAATTGCTTTCTTCGCAACGCAAAAGTTAAGATTTCAGATTATCAGGGCGCAATG 1062
Qy 338 -----GluProArg 340
Db 1063 CTACGAGAAAAACAGAACCGAAA 1086

RESULT 13.
US-09-769-744A-93
; Sequence 93, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamifly, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769, 744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 93
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-93

Alignment Scores:
Pred. No.: 8.3e-07 Length: 1200
Score: 151.50 Matches: 98
Percent Similarity: 41.58% Conservative: 65
Best Local Similarity: 25.00% Mismatches: 165
Query Match: 7.50% Indels: 65
DB: 12 Gaps: 16

US-09-889-756a-2 (1-412) x US-09-769-744A-93 (1-1200)

Qy 12 AlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyIleAspAla 31
Db 43 GCAGCAATCGCTGCTGCGAGTAGTGTATG-----GGTCTGGGGGCAATTTTA 93
Qy 32 -----AlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaProValValGlyVal 49
Db 94 CTTTGAAGAACCTTCTCTCACTGCTCTTAAGAGTAGAGCT-----ACTCATCTT 144
Qy 50 ValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSer 69
Db 145 GTTGTCGCCAAGAGAGAGAGCGCTGCTCTCTGTTTATTTGTCAGGACAGTAACGCA 204
Qy 70 LeuArgThrAlaAspValArgAlaGlnValGly--GlyIleIleGlnLysArgLeuPhe 88
Db 205 AAAAATGAACAAATATGTTTATTTTTCATGCTAGTAAGGTCATTTAGTAATCCTTGT 264
Qy 89 GlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyr 108
Db 265 TCTGTGGCGATTAAGTACGCGAAGCGGCTTTAGTCAAGTAACGTAGTTCAGAACGC 324
Qy 109 GluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAla----- 122

```



```

Db 7419367 GGCCTTCGATCGGGGGTGGTAGCGGACGTACCTCACTATCCGCTGACCCCGGG 7419308
Qy 202 euasnaalaglyAspThrValIleuAlaThrIleArgIleThrAspProwet----- 219
Db 7419307 CGAACCGTCACCCGGGGTTCACCGGGCCACGGGTGAGCGGGCTACAGCCGGTTGCGCGTG 7419248
Qy 220 -----TyrV 221
Db 7419247 TGACACCGCGCAGGACCGCCCGCTTGTGGGGACCTCGTCCTCGGCGACCGGG 7419188
Qy 221 alaenValThrIleSerAlaSerGluValMetIleuArgIleGlnIleAlaGluIle 241
Db 7419187 TCCGAGTCCGCGACGACGATCCCGCCCGCTGCACTGACGCGCGTCCGCGACGAC 7419128
Qy 241 yslleuValAlaAlaAspGlyValIleAlaValGlyIleIlePhe-----Aspasp 258
Db 7419127 GCCCGCGGATGGGATGGCGGTGTCCGAGTCCGCCCGGAAGTGCAGGTACCCACGAC 7419068
Qy 258 lYThrValTyrProGluIleGlyIleuLeuPheAlaAspProVal----- 273
Db 7419067 CCGCGGTACAGCCCGCCCGGAGCGGCTCCAGTTCGTGATGATCTGACAGGGCGCGGGG 7419008
Qy 274 -----ValAsnGluSerThrGlyIleIleThrIleuArgAlaAlaVal----- 287
Db 7419007 TTGGGCGCGCGGAGAGGGTCCGCGCGGAGACGCGCTGACACGCTGACAGCGCGTA 7418948
Qy 288 -----ProAsnAspGlnAsnIleLeuMetProGly-----LeuTyrV 300
Db 7418947 CGCCCCCGCGGACCGC-----CCCGGTACCGGTGACACAGATGTGCATACG 7418900
Qy 300 alaArgValIleuMetAspGlnValAlaValAspAsnAlaPheValValProGlnIleAlaVal 320
Db 7418899 TGGCAGTAGCGCTCGATGCTGATGAGTGCAGAC----- 7418865
Qy 320 alThrArgIleAlaLysAspThrValMetIleValAsnAlaGlnIleGlyIleMetGluProA 340
Db 7418864 -----CTCGACGGAACCGGGCTCGACACCGCG 7418837
Qy 340 rglIuValThrValAlaGlnIleGlnIleThrAsnTrpIleValThrSerGlyIleuLysA 360
Db 7418836 CCCAGATC---GTTCGCGCCGAGTGCAGACGATCAGTGTCTGGCGCTCTTGGGG 7418780
Qy 360 spglAspLysValValAlaGluIleIleSerIleAlaGlyIleThrGlyAlaLysLys 379
Db 7418779 TCGGCGCAGACGCTGTCCGCGAGGGCTGTCTCTCGCGGGGTGCGCGCGCGCGCG 7418720
Qy 380 ValThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyVal 399
Db 7418719 GTGCCCGCGATGGGGTGGAC-----ATCGCTGCCCGTCTCGACCTTG 7418675
Qy 400 GlnThrAlaSer 403
Db 7418674 ACAGGGGCTCG 7418663

RESULT 15
US-09-741-669-242
; Sequence 242, Application US/09741669
; Patent No. US2002022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zykkind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741.669
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: PaasSEQ for Windows Version 4.0
; SEQ ID NO 242
; LENGTH: 1068

```

```

; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1068)
US-09-741-669-242

Alignment Scores:
Pred. No.: 1 79e-05 Length: 1068
Score: 138.50 Matches: 52
Percent Similarity: 42.42% Conservative: 32
Best Local Similarity: 26.26% Mismatches: 55
Query Match: 6.86% Indels: 59
DB: Gaps: 6

US-09-889-756a-2 (1-412) x US-09-741-669-242 (1-1068)

Qy 65 GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
Db 121 GGGCGCATTTGAAACGACGAA---GTGATATTGCCAGCAAAATTGCCGGCGTATCGAC 177
Qy 85 LysArgLeuPheGlnIleGlySerIleValArgAlaGlyIleProLeuTyrGlnIleAsp 104
Db 178 ACAATTCTGTGAAGAAGGCAAGTTGTTCGCGAAGTGAAGTGTGCGGAGATGAT 237
Qy 105 SerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 238 ACTCGCGTGTTCGAGAAACGACGATCGAAGCCATCGCGCAATCAAGAGCACAAAGC 297
Qy 125 ThrLeuAlaLysAlaAspAlaAspLeu----- 133
Db 298 GCCGTTCGTCGCGCAGCGCTTGTTCGAGCAACGAAAGCAAACTCGTCGCGACAG 357
Qy 134 -----AlaArgTyr 136
Db 358 TCGCTGTTAATCAACGCCGACGAACTGACTCCGTACGAAACGTCATACGCTTCC 417
Qy 137 LysProLeuValAlaAlaGlnAlaValSerArgGlnIleTyrAsp----- 151
Db 418 GTTCACTGCGCCCAAGAGGGGCTATTTCGCCCAACGCTGATGACATCCGCCGCC 477
Qy 152 -----AlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAla 169
Db 478 GCTGACGCGCCGACGCTGCTGGAATGCGGAAAGCTAGGTATCGCTTCTAAAGC 537
Qy 170 AlaIleSerAlaGlyIleAsn----- 177
Db 538 GCTATGAAAGCGGACGACCAATATTCAGGCGCAACCCGCTCGAAGCGCACAA 597
Qy 178 -----LeuAsnArgSerArgIleThrAlaProIleSer 188
Db 598 GCCACTGACGCGGCTTCCCGCATATGATGACGACGCACTGAAGCCCGCGTGAC 657
Qy 189 GlyPheIleGlyIleSerLysValSerGlu---GlyThrLeuLeuAsnAlaGly 205
Db 658 GGAAGCGGTG---CAGTATCGGGTGTGCGAGCCAGGGAAGTGTGGCGGAGGC 708

RESULT 16
US-09-855-754-1/c
; Sequence 1, Application US/0985754
; Publication No. US20020192237A1
; GENERAL INFORMATION:
; APPLICANT: BOURSAUX-BUDE, CAROLINE
; APPLICANT: GUIRO-MACLOUF, NICOLE
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING POLYMORPHISMS OF THE REPEATED
; TITLE OF INVENTION: REGIONS OF PERTACTIN IN BORDETELLA PERTUSSIS,
; TITLE OF INVENTION: BORDETELLA PARABERTUSSIS, AND BORDETELLA
; TITLE OF INVENTION: BRONCHISEPTICA, THEIR USE IN DIAGNOSTICS, AND IN
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS
; FILE REFERENCE: 03495-0206-00000
; CURRENT APPLICATION NUMBER: US/09/855, 754
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/206,969

```


Db 2687 CCCTGCGCCGGAGATCCGCGCTACGCCGCCGAGACACCGCGCCGCAACTCCGAGCGCGCG 2628
 Qy 71 ArgThralaspValArgAlaGlnValGlyGlyIleGlnLysArgLeuPheGlnGlu 90
 Db 2627 AAGACGAGATGCGCCCGTCCGCGCGCTCCAGCGGTGTCCACCGCG-----TCGCC 2586
 Qy 91 GlySerThrValArgAlaGlyGlnProLeuThrGlnIleAspSerSerThrTyrrGluAla 110
 Db 2579 GACACCTTCGACGCTCCAGCATCTCGGTGTTCCTCGCTGCTGCTGCAAGTC----- 2523
 Qy 111 AsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaIysAlaAsp 130
 Db 2528 -----GCCGCTTCAGCGCCAGCGCAGCATGAGCGCGGGGGCGCGCGCTCGCGCGTTC 2475
 Qy 131 AlaAspLeuAlaArgTrpLysProLeuValAlaAlaGluAlaValSerArgGlnGluTrp 150
 Db 2474 AGCTGATGTCGTGTAACGACGCCGATGTCCTCGCTGACAGCGCGCAGCAGGATTCGTGTA 2415
 Qy 151 AspAlaAlaValThrAlaLysArg-----SerAlaGlu 161
 Db 2414 GACCCCGGCTCACCCGTCGCGCGCGCGCTCCAGTCCAGAACGCAAGCGCGCCAC 2355
 Qy 162 AlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAla-----Gly 175
 Db 2354 GCGAGTGGAGAGGCGCAGCTCCGCGAGCGCGCGGGCGAGCGCAGCGCGCGGTGGCG 2295
 Qy 176 IleAsnLeuAsnArgSerArg-----IleThrAlaProIleSerGlyPhe 191
 Db 2294 GCCAGTTCGCGCGCGCGCGCGAGCAGCAGCGGTCCACCGCGCGATTAAGCTGGCGCC 2235
 Qy 192 GlyIleSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThr---ThrValLeu 210
 Db 2234 ACCTGCTCGCGCGCGCGGAGCCCGTGTTCGTGACGTGAAGAGAACCCGACCTGTCT 2175
 Qy 211 AlaThrIleArgGlnThrAsnProMetTrpValAsnValThrGlnSerLysGluVal 230
 Db 2174 TCGTCCGCGCGGTGTTCGAGGCCCTCTGTACGCGAACGA----- 2137
 Qy 231 MetLysLeuArgArgGlnIleAlaGlnIleLysLeuLeuAlaAlaAspGlyValIleAla 250
 Db 2136 ---GCGGTACACAGCGCCGTGCGAACG----- 2110
 Qy 251 ValGlyIleLysPheAspAspGlyThrValTyrrProGluLysGlyArgLeuLeuPheAla 270
 Db 2109 -----GAACAGGGCGGTGCCCTCGAAGTACAGATTCGTGGCGGTCC 2065
 Qy 271 AspProValValAsnGlnSerThrGlyGlnIleThrLeuAlaAlaValProAsnAsp 290
 Db 2064 GATCACGTCGTAGTCGAGAGGCGCGGCGAGGTCCCGCTC----- 2028
 Qy 291 GlnAsnIleLeuMetProGlyLysLeuTrpAlaArgValLeuMetAspGlnValAlaValAsp 310
 Db 2025 -----GTCCCGGTACAGGTGTG-----CGGAGCGCGCCCGC 1997
 Qy 311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIle 330
 Db 1992 GGGGAGACGTGCGCGGACAGGACCGGGCCAGCGCGGTGCCCGCGCTGCGCCCGTGT 1933
 Qy 331 ValAsnAlaGlnGlyLysMetGluProArgGluValThrValAla----- 345
 Db 1932 CGCGGTCCACAGGGCGGC-----CGGAGGTCTCGCGTGTCCACCGCGTACGGGTA 1882
 Qy 346 -----GlnGlnGlnGly 349
 Db 1881 CGCCGACACCAAGACCAAGGC 1861

RESULT 21
US-10-227-353-5/c
; Sequence 5, Application US/10227353
; Publication No. US20030087383A1
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A

```

1 TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
2 YEAST
3
4 NUMBER OF SEQUENCES: 17
5
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Milten, White, Zelano & Branigan, P.C.
8 STREET: 2200 Clarendon Blvd., Suite 1400
9 CITY: ARLINGTON
10 STATE: VA
11 COUNTRY: USA
12 ZIP: 22201
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patentin Release #1.0, Version #1.30
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/10/227,353
22 FILING DATE: 26-Aug-2002
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US/08/460,269C
26 FILING DATE: 02-Jun-1995
27
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Lebovitz, Richard M.
30 REGISTRATION NUMBER: 37,067
31 REFERENCE/DOCKET NUMBER: Popov-2
32
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: (703) 243-6333
35 TELEFAX: (703) 243-6410
36
37 INFORMATION FOR SEQ ID NO: 5:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 3000 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: single
42 TOPOLOGY: linear
43
44 MOLECULE TYPE: DNA
45
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 145..2910
49
50 SEQUENCE DESCRIPTION: SEQ ID NO: 5:
51
52 US-10-227-353-5
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
```



```

QY 280 GnuILetHrLeuArGAlaAlaValProSnpAspGlnAnIleLeuMetProGlyYeuYr 299
Db 1914 -----CTGAAAACAAACTCAGCGACGTCCGACCACTATCAC 1876
QY 300 ValArGValLleuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAla 319
Db 1875 GTTCAGCGCG---CGGAAGAAGTTAGCAAAAGCCATTGCGGAAGTGAACCAAGCGTTCCG 1819
QY 320 ValThrArGValAlaLysAspPhyValMetIleValAn-----AlaGlnGly 335
Db 1818 GTGATCCGTGACGGGAAGACTCTCTCCACAGCGCTCAACTTACCGCGCCGATGCGAGGC 1759
QY 336 GlyMetGluProArgGluValThr 343
Db 1758 ATTGTCAGGATATTGATGTCCACC 1735

RESULT 24
US-10-114-170-39
; Sequence 39, Application US/10114170
; Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blatcner, Frederick R.
Perna, Nicole T.
Burland, Valerie
Plunkett, Guy
Melch, Rod
TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESS: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION DATA: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296,95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 25165
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-114-170-39

Alignment Scores:
Pred. No.: 0.01 Length: 25165
Score: 130.50 Matches: 61
Percent Similarity: 40.28% Conservative: 55
Best Local Similarity: 21.18% Mismatches: 85
Query Match: 6.46% Indels: 87

```

DB:	14	Gaps:	9
US-09-889-756A-2 (1-412) x US-10-114-170-39 (1-25165)			
QY	60	ThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal	79
Db	24240	ACGGTAGAACCTTCGGGCGCGAAGAGGTGTGCATCGCTGGAA-----	24286
QY	80	GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnPro	99
Db	24285	GGCGGAATTCGTATCATCTCGACGTGAAGATTGGGATATTGTCAACAGGCGCAGCGG	24344
QY	100	LeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeu	119
Db	24345	CTGGCGCAGCTTACCCGCCCAAAACGGAATTCGAGCTCCAGGAACGATGTGGCGCTG	24400
QY	120	AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeu	139
Db	24405	TACCGCGCGCTGGGAACCTCGGGCGGCTCGGGCGGAAGTAGCAAT---AAACGGCTG	24460
QY	140	ValAlaAlaGluAlaValaSerArgGlnGluTyrAspAlaAlaValaThrAlaLysArgSer	159
Db	24462	GTCTTCCCTCAGCAACTGATATAA-----TTCCCGCAACTTATCAGATCGC-----	24508
QY	160	AlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerLysAlaGlyIleAsnLeuAsn	179
Db	24507	-----GAAACGGCGGTGTGACCAACTCGC	24530
QY	180	ArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGly	199
Db	24531	CGGCAC-----CGGTGACAAAGCCACACCGG	24566
QY	200	ThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMet	219
Db	24561	-----CTGACGACGGGAGATTACCTCGTCAATCGGCACTGGCGATGACCCAGCGCTG	24614
QY	220	TyrValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlu	239
Db	24615	GTAACAAGCGGGCGACGACGACGCTCGAAGTACTCGCTCCACAGCCAGGCCAATGAA	24674
QY	240	GlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThr	259
Db	24674	-----	24674
QY	260	ValTyrProGluLysGlyArgLeuPheAlaAspProValValAsnGluSerThrGly	279
Db	24674	-----	24674
QY	280	GlnIleThrLeuArgAlaAlaValaProAsnAspGlnAsnIleLeuMetProGlyLeuTyr	299
Db	24675	-----CTGGAAAAAACAATCGACGACGCTGGCAGCAGCAATACTAC	24713
QY	300	ValArgValLeuMetAspGlnValaAlaValaLysAsnAlaPheValValProGlnGlnAla	319
Db	24714	GTTCAAGCGC---CGCAGAGAGTTAGCAAAAGCCAAATGCGGAAGTGAAACCAAGCGCTTG	24770
QY	320	ValThrArgGlyAlaLysAspThrThrValMetIleValAsn-----AlaGlnGly	335
Db	24771	GTCATCCGTGACGAGGAAGACTCTCCACAGGCTCAACTTTACCGCGCGGTGCGAGGC	24830
QY	336	GlyMetGluProArgGluValaThr	343
Db	24831	ATTGTTCAAGATATTGATGTCTCAC	24854
RESULT 25			
US-09-938-842A-176			
/ Sequence 176, Application US/09938842A			
/ Patent No. US20020160378A1			
/ GENERAL INFORMATION:			
/ APPLICANT: Harper, Jeff			
/ APPLICANT: Kreps, Joel			
/ APPLICANT: Wang, Xun			
/ APPLICANT: Zhu, Tong			

```

# TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
# TITLE OF INVENTION: SAME AND METHODS OF USE
# FILE REFERENCE: SCDIP1300-3
# CURRENT APPLICATION NUMBER: US/09/938, 842A
# CURRENT FILING DATE: 2001-08-24
# PRIOR APPLICATION NUMBER: US 60/227, 866
# PRIOR FILING DATE: 2000-08-24
# PRIOR APPLICATION NUMBER: US 60/264, 647
# PRIOR FILING DATE: 2001-01-16
# PRIOR APPLICATION NUMBER: US 60/300, 111
# PRIOR FILING DATE: 2001-06-22
# NUMBER OF SEQ ID NOS: 5379
# SEQ ID NO 176
# LENGTH: 1539
# TYPE: DNA
# ORGANISM: Arabidopsis thaliana
# US-09-938-842A-176

Alignment Scores:
Pred. No.: 0.000277 Length: 1539
Score: 129.50 Matches: 95
Percent Similarity: 35.27% Conservative: 97
Best Local Similarity: 22.04% Mismatches: 168
Query Match: 6.41% Indels: 111
DB: 10 Gaps: 18

US-09-889-756A-2 (1-412) x US-09-938-842A-176 (1-1539)
Qy 38 AlaGlyArgGluAlaProAlaProValAlaGly-----ValValThrValHis 53
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 GCCTGTGACTGACGCTTCAGCGGCTCTCTCCGCGACGAGAGAGATGTTCTCACATCCAC 228
    54 ProGlnThrValAlaLeu-----ThrValGluLeuPro--GlyArgLeu 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 229 GCGCGGATGCTGCCACTCATTTGACAAATCTTACAGCGGTAGGCTCGCTCGCGGATCTC 288
    68 GluSerLeuArg-----ThrAlaAspValArgAlaGlnValGlyGly 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 GAGATTCTCCGCTCTTTGTTCAAGAGATATCACCGCTCGCGGTTTCGCTCGCGGTGAC 348
    82 IleIleGlnLysArgLeu-----PheGlnGluGlySerTyr 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 349 GAGATCCAAATGGCCGTTACTTAAGACGACCCCGCTTAAGTCGATGATCATTCAC 408
    94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsn--Leu 112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 TTCCTCTCGCTCCGACCCGTTAAAGAAATCCAGATCATTCGATCAACGAAACA 468
    113 GluSerAlaThrArgAlaGlnLeuAlaThrAlaGlnAlaThr-----Leu 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 469 GAGAAACGATGTTGAACCTACGATTAACAAATGGCTTGAAAGGTCAAGAACCGATGCTT 528
    127 AlaLysAlaAspAlaAspLeuAlaArgTyrTyrProLeuValAlaAlaGluAlaValSer 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 GAGAGGCTAGACACAGATCTTAGCTGATTACAGTACTTTTCAACGCGGAGGAGAAACAGAA 588
    147 ArgGlnGluTyrAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLysAla 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 589 GAAGAGAAACGTTTGAGATTGACGCGCGCGAAGGAGACTTCAACGCGAGACCTGAAGGGA 648
    167 AlaGlnAlaAlaIleLysSerAlaGlyLysLeuAsnLeuAsnArgSerArgIleThrAlaPro 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AAGAGGAAAAAGGCTGTCGAGAGACAGTGTACGCGCTTATTGACGACACTGCTCCGAAT 708
    187 IleSerGlyPheIleGlyGlnSerLys-----ValSerGluGlyThr----- 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GTTGAGGATTTACAGTGGCGCTTGCAACGAACTGATCGTCTGCTTCTGCTCAATTGATT 768
    201 -----LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle----- 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 769 AAAGGAGATCTTATGCTGTGGGATCTCAACATGATGATGATGCGGAGAAATGATTTTC 828
    214 -----ArgGlnThrAsnProMetTyrValAsn 222

```


Query Match: 6.32% Indels: 67
DB: 11 Gaps: 6
US-09-889-756a-2 (1-412) x US-09-988-067B-51 (1-1116)

QY 63 LeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIle 82
Db 201 TTGCAGAGGTTTGGAGCC---AGAGATATACAGCGTAGCTCCAAAGTCCCTGGCCG 257
QY 83 IleGlnLysArgLeuPheGlnGlySerIleValArgAlaGlnProLeuThrGln 102
Db 258 ATTGAAAAGGTTTGTAAAGGCGATCATTAAGGCGGATTTGGTTTTCG 317
QY 103 IleAspSerSerThrIleGlnAlaAsnLeuGluSerIleAlaGlnAla 117
Db 318 ATTCTAGCCCTGAATTAAGAACCAACTCCCTCAGCTGAAGCCGGCATTAAGCCGCT 377
QY 117 ----- 117
Db 378 AAAGCGCTTAGCATGAAGTCAAAAGAGCTCAAGAGACGAACGATTAATTTCTGCGAGA 437
QY 118 -----GlnLeuAlaThrAlaGlnAlaThrLeuAla----- 127
Db 438 GACGTTTGGCAAGACGCCAAATCCCAAGCCACTTTAGCCAAAGAGACTTAATAGCGCGTT 497
QY 128 -----LysAlaAspAlaAspLeuAlaArg 135
Db 498 CAAAGTTTGTATGATTAATGCGCGCTGCGAGCTTGCAAAAGCCGATGAAGCCTATGCGGCT 557
QY 136 TyrLysProLeuValAlaAlaGlnAlaValSerArgGlnIleIleValAlaAlaVal 154
Db 558 TATGAAGCAGCTAAATACAGAGAGCGCGCTTCCAAAGTATTAATAGCTTACGG 617
QY 155 -----ThrAlaLysArgSerAlaGlnAlaGlyValValAlaGlnAla 170
Db 618 GCGGCGAGCTCTGAAGATTAATGCGCTGAGCTTAAGAGAGAGCGCGCTTTAGCGCA 677
QY 171 IleLysSerIleGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 678 GTGAATGAAGTGAAGCTTATTAATAAGACGTCAAGAGAGAGAGAGAGAGAGAGAGAG 737
QY 191 IleGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 738 GTGAGTAAAGCTTAAAGCGGTGCGAGCTTAAAGCTTAAAGCGGTGCTTCTGCTTTTA 797
QY 211 AlaThrIleArgGlnThrAsnProMetIleValAlaValThrGlnSerAlaSerGlnVal 230
Db 798 ATGATAGATTTAAAGATGTTGTTAAATAATCAAGCGTGAAGAGATTTTGAACGAG 857
QY 231 MetLysLeuArgArgGlnIleAlaGlnGly----- 240
Db 858 TTTAAAGTGGTAAAGAAATTT---CAAGGCTATATCCCGCGCTTCAAAAAAGCAGCAA 914
QY 241 -----LysLeuLeuAlaAlaAspIleValIleAlaValGlyIleLysPheAsp 257
Db 915 TTCAAGGCTCAATATTGAGCGTGAATGGGGATTTTGGAGCTTGAAGAGCAGATATAT 974
QY 258 GlyThrValIleProGlnLysGlyArgLeuLeuPheAlaAspProVal 273
Db 975 TCCAACTTACGACATGAAGAGCTATGATGAAGGAGCAATACCTTA 1022

RESULT 28
US-09-793-306-145
Sequence 145, Application US/09793306
Patent No. US20020098200A1
GENERAL INFORMATION:
APPLICANT: Campos-Neto, Antonio
APPLICANT: Skeiky, Yasir
APPLICANT: Ovendale, Pamela
APPLICANT: Jen, Shyan
APPLICANT: Jodes, Michael
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

TITLE OF INVENTION: of Tuberculosis
FILE REFERENCE: 014058-008740US
CURRENT APPLICATION NUMBER: US/09/793,306
CURRENT FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 60/185,037
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/223,828
PRIOR FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 164
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 145
LENGTH: 1794
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
NAME/KEY: CDS
LOCATION: (1)..(1794)
OTHER INFORMATION: mTTC#3-His (MTB57)
US-09-793-306-145

Alignment Scores:
Pred. No.: 0.000923 Length: 1794
Score: 125.50 Matches: 91
Percent Similarity: 31.54% Conservative: 50
Best Local Similarity: 20.36% Mismatches: 164
Query Match: 6.22% Indels: 142
DB: Gaps: 14

US-09-889-756a-2 (1-412) x US-09-793-306-145 (1-1794)

QY 1 MetValPheThrAlaPheLysAlaMetArgAlaAlaAlaAlaAlaValAlaLeu 20
Db 175 TTGCGGCTAGCTCTGCGAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 234
QY 21 ValLeuSerSerCysGlyLysGlyAspAlaAlaGlnIleGlyGlnProAlaGlyArg 40
Db 235 TAT-----GGGGGTGTTGGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279
QY 41 GlnAlaProAlaProValValGly-----ValValThrValHis 53
Db 280 TCGGCTCAGGCAAGCGGTCGCGCAATGCTTGAAGCGCGCGCGCGCGCGCGCGCAT 339
QY 54 ProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuSerLeuArgThrAla 73
Db 340 CCGATGCTGTCGGCGCCAAAC-----CGTAATGCG 369
QY 74 AspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlySerTyr 93
Db 370 TTGTGCGAG-----TTGGTGTTCGAATCTGTTGGGCAAGATGCG--- 411
QY 94 ValArgAlaGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGln 113
Db 411 ----- 411
QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeu 133
Db 411 ----- 411
QY 134 AlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGlnIleTyrAspAlaAla 153
Db 412 -----CCGGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 462
QY 154 ValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173
Db 463 GTGCGCGGATGTCGCTATCACGCGGCGGATCGCGCGCGCGCGCGCGCGCTGTCGCG 522
QY 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaPro-----IleSerGlyPheIle 191
Db 523 TGTCAATTTGCTGCGAGCGCTTCCAGCTGCGCATTCGCGCGCGCGCGCGCGCAT 582
QY 192 GlyLysSerLysValSerGlnGlyThrLeu-----LeuAsnAlaGlyAsp-ThrThrVa 209

```

Db      583 GGCCTCGGCAACATCGGCTCGGAACTCGGCGCGGGAACCCGGTACTACATCTG 642
Qy      209 lLeuAlaThrIleArgInThr-----AspProMetTy 220
Db      643 GGCACGGCAAAATTCGGGCAACCCCAAGTAGTAGCGGAACTCCGGCAAGCCCAATGTG 702
Qy      220 r-----ValAsnValThrGln---Se 226
Db      703 GGCACGGCAAAATGACGTCGCCAAGATTGGGACGGAAATATGCGCAACCCATCTC 762
Qy      226 rAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLysLeuLeu-----Al 244
Db      763 GGCACGGCAAAATGTCGATCTCAATCTGGGACGGCAACCCGAGCTTTGGAAACCTC 822
Qy      244 aAlaAspGlyValIleAlaValGlyIleLysPheAspGlyThrValTyrProGluLys 264
Db      823 GGCACGGCAAACTTTGGCAGTGGG----- 847
Qy      264 sGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuAr 284
Db      848 -----ACCTGGGCAAGTGGAAACCCGGAAGTACCAACTTC 882
Qy      284 gAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMe 304
Db      883 GCGCGCG----- 889
Qy      304 tAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAl 324
Db      890 -----GAAATCTCGGTTCTTCAACTGGGCAAGTGGAAACATCGCTCTCCACATCGGT 945
Qy      324 aLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrVa 344
Db      946 TTCGGAACAACACGGGCAATACCTCGCTCGGCAACATGGCAACAACAATCGGT 1005
Qy      344 lAlaGlnGlnGlnGlyThrAsnThrPileValThrSerGlyLeuLysAspGlyAspLysVa 364
Db      1006 TTTGGGCTCACCGGCAACATCTGGTGCGCATTTGGCGCTGAACCTCGGCATCGGAGAT 1065
Qy      364 lValValGluGlyLysSerIleAlaGlyIleThrGlyAlaLysValThrProLysG 384
Db      1066 CTAGGTTTCGGGAACCTCGGTACCAACAACATCGTTCTTCAACTCGGCAACAACAAC 1125
Qy      384 uTrpAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGlnThrAlaSerG 404
Db      1126 GTGGGCTTCTTCATTCG-----GCAACAACAATTCGCTTTGGAACGCGGCGAC 1179
Qy      404 uAlaLysThrAlaSerGlu 410
Db      1180 ATCAACACGGGCTTCGGA 1198

```

RESULT 29

```

US-09-847-513A-1
; Sequence 1, Application US/09847513A
; Publication No. US20030104375A1
; GENERAL INFORMATION:
; APPLICANT: MBARI
; APPLICANT: DeLong, Edward
; APPLICANT: Beja, Oded
; TITLE OF INVENTION: Light-driven energy generation using proteorhodopsin
; FILE REFERENCE: MEA-101
; CURRENT APPLICATION NUMBER: US/09/847,513A
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 60/201,602
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent version 3.0
; SEQ ID NO 1
; LENGTH: 105184
; TYPE: DNA
; ORGANISM: Naturally occurring gamma proteobacterium
; FEATURE:
; NAME/KEY: gene

```

```

; LOCATION: (50866)..(51615)
; OTHER INFORMATION: Proteorhodopsin gene sequence.
; NAME/KEY: misc feature
; LOCATION: (1593)..(2807)
; OTHER INFORMATION: Predicted threonine dehydratase. Contains 'n' at position 2753.
; PUBLICATION INFORMATION:
; AUTHORS: Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P.,
; AUTHORS: Jovanovich, S.B., Gates, C.M., Feldman, R.A., DeLong, E.F.
; TITLE: Bacterial rhodopsin: evidence for a new type of phototrophy in the sea
; JOURNAL: Science
; VOLUME: 289
; ISSUE: 5486
; PAGES: 1902-1906
; DATE: 2000-09-15
; DATABASE ACCESSION NUMBER: AF279106
; DATABASE ENTRY DATE: 2000-06-15
; RELEVANT RESIDUES: (50866)..(51615)
US-09-847-513A-1

```

Alignment Scores:

Pred. No.:	0.281	Length:	105184
Score:	125.00	Matches:	72
Percent Similarity:	38.01%	Conservative:	50
Best Local Similarity:	22.43%	Mismatches:	147
Query Match:	6.19%	Indels:	52
DB:	11	Gaps:	9

US-09-889-756a-2 (1-412) x US-09-847-513A-1 (1-105184)

```

Qy      58 AlauLeuThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAla 77
Db      88348 GCGTAACTAGAGAAATTCCTCGGAAGCTCATTCACACCAAGCTTAAGCTAGCATTT 88407
Qy      78 GluValGlyGlyIleIleGlnLysArgLeuPheGluGluGlySerTyrValArgAlaGly 97
Db      88408 GAAATACCTCGAAAGATAACTATTATGTTGATATCGAGATGAGTCACTTTAGGG 88467
Qy      98 GluProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAla 117
Db      88468 GATGAACCTTCCTCATTTGATGAT-----ACAGAGCTTTTAGCA 88506
Qy      118 GluLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLys 137
Db      88507 CAATTAATCAATCAACAAGCTAATTTGATTAGCTGGAACAAGTACTAGCAAGATATATA 88566
Qy      138 ProLeuValAlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLys 157
Db      88567 AATCTCAGACAGATGCGCATATTTCTATTCAAGATCTTGAT----- 88608
Qy      158 ArgSerAlaGluAlaGlyValLysAlaAlaAlaAlaIleLysSerAlaGlyIleAsn 177
Db      88609 ---AAGCTGAATCTGATCAATCAATAGTAGCTAGCTGACAGTATGATTATTAAGAGTTAA 88665
Qy      178 LeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyLysSerLysValSer 197
Db      88666 TTGAGCAACAACTAAGTACTAGCTCTTTAATAGAGTATTCAAAATGATTCTTAGAT 88725
Qy      198 GluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsn 217
Db      88726 ACAGATCAAGTATCAATGACAGGT---GTCCAAGTTTGAATAATTTAGGCTCTAGCAAT 88782
Qy      218 -----Promet---TyrValAsnValThrLysSerAlaSerGlu 229
Db      88783 GTTGAAGCAAGAAATTTCTATTCCAAATGAACATAACAGGTTAAGTTGAGAGTAGAG 88842
Qy      230 ValMetLysLeuArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAspGlyValIle 249
Db      88842 ----- 88842
Qy      250 AlaValGlyIleLysPheAspGlyThrValTyrProGluLysGlyArgLeuLeuPhe 269
Db      88843 -----TATCAATTTGATATCAGAGAGATA---TCTACAAAGCTACGTTAGAGAGA 88890

```

```

QY 270 AlaaspProValValAsnGluSerThrGlyGlnIleThrLeuArgAlaValProAsn 289
Db 88891 CTGGCTCCCATG-----TCAACCGAGGCTCCAAATAATAGGTAGCATTTTGA 88941
QY 290 AsgGlnAsnIleLeuMetProGlyLeuTyrrValArgValIleLeuMetAspGlnValAlaVal 309
Db 88942 TTTGATACCTTTTAAATCCAGATCAATAGTAAAGCTCAACATAGCATCACTGAGAA 89001
QY 310 AspaenAlaPheValAlaProGlnGlnAlaValThrArgGlyAlaIleAsp-----Thr 327
Db 89002 GCAAAAGGTACATGGTTCATTAAGTCACTGCTCCAGTCCGAAACAAGTATTTGGGCC 89061
QY 328 ValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGln 347
Db 89062 AATATATACATTAACGACCAACAAAGTATGTTAGAGATCTGTGATGTTATTTATTTT 89121
QY 348 GlnGlyThrArgThrPleValIleThrSerGlyLeuLysAspGlyAspLysValValGlu 367
Db 89122 GAAGAAGAAATATGCTTTTGTACGCGAACACTTAAATATGATTTAGTAAATTTAGGC 89181
QY 368 G1Y 368
Db 89182 GGA 89184

RESULT 30
US-08-945-038-3
Sequence 3, Application US/08945038
Publication No. US20020146423A1
GENERAL INFORMATION:
APPLICANT: Doidge, Christopher V.
APPLICANT: Lee, Adrian
APPLICANT: Radcliffe, Fiona J.
APPLICANT: Hocking, Diana M.
APPLICANT: Webb, Elizabeth A.
TITLE OF INVENTION: PROTECTIVE HELICOBACTER ANTIGENS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,038
FILING DATE: 23-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00225
FILING DATE: 19-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN2575
FILING DATE: 21-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN3931
FILING DATE: 03-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN7565
FILING DATE: 16-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 017227/0133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 base pairs

```

```

/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ ORIGINAL SOURCE:
/ ORGANISM: helicobacter pylori
/ IMMEDIATE SOURCE:
/ CLONE: clone E2.5
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..987
/ US-08-945-038-3

Alignment Scores:
Pred. No.: 0.000859 Length: 990
Score: 122.50 Matches: 63
Percent Similarity: 36.23% Conservative: 37
Best Local Similarity: 22.83% Mismatches: 109
Query Match: 6.07% Indels: 67
DB: Gaps: 6

US-09-889-756a-2 (1-412) x US-08-945-038-3 (1-990)
QY 63 LeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIle 82
Db 112 TTGCAGAGATTTTGTGAAGCC--AGAAATACAGCGTGATTCCAAAGTCCCTGGCCG 168
QY 83 IleGlnLysArgLeuPheGlnGlySerTyrrValArgAlaGlyGlnProLeuTyrrGln 102
Db 169 ATTGAAGAAGTGTGTTTAAAGGCGATCGCATTAAGGCGATTGGTTTAC 228
QY 103 IleAspSerThrTyrrGluAlaAsnLeuGluSerAlaArgAla----- 117
Db 229 ATTCTAGCCCTGATATAGAACCAAGCTCGTCAACGTGAAGCGCGCATTAAGCGGCT 288
QY 117 ----- 117
Db 289 AAAGCGTTAGCGATGAGTCAAAAGAGCTCAAGAGCAAGCAATTCACAGA 348
QY 118 -----GlnLeuAlaThrAlaGlnAlaThrLeuAla----- 127
Db 349 GACGTTGGCAAGCGGCAAAATCTCAACCACTTACCAAGAGACTTATAGCGGCTT 408
QY 128 -----LysAlaAspAlaAspLeuAlaArg 135
Db 409 CAAGTTGTATGATATAGCGGTGGCGAGCTTGCAAAAGCGCATGAAGCTATAGCGGCT 468
QY 136 TyrrLysProLeuValAlaAlaGluAlaValSerArgGlnGlyTyrrAspAlaAlaVal--- 154
Db 469 TATGAAAGCACTAAATACAAAGAGCGCGGCTTACCAAAAGTATTAATGCGCTTTAGGG 528
QY 155 -----ThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAla 170
Db 529 GGGCGAGCTCTGAAGATGAGATTGCGCTAAGCTAAAGAGCGGCGCTTTAGGGCAA 588
QY 171 IleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 589 GTGAATGAAGTGAAGTCTTTATTAAGATGTCAAGACGACGCCAATTGATGGGAA 648
QY 191 IleGlyGlnSerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 649 GTGAGTATGTGCTTTAAGCGGTGGCGAGCTTACCCCTAAGGCGCTTCCGTGTGCTC 708
QY 211 AlaThrIleArgGlnThrAsnProMetTyrrValAsnValThrGlnSerAlaSerGluVal 230
Db 709 ATGATGATTTAAAGATAGTGTGTTAAATAATCAGCGCTGCTGAAGATTTGAACAT 768
QY 231 MetLysLeuArgArgGlnIleAlaGlyGly----- 240
Db 769 TTTAAAGTGGTAAGAAATTT--GAGGTTATATCCCGCGTTGAAGAAAGCGCGAAA 825
QY 241 -----LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAsp 257

```

Db 826 TTCAGGCTCAATATTGAGCGCTGATGGGGGATTTTGGCACTTGGAAGCAAGCAATPAT 885
 Qy 258 GlyThrValTyrProGluIleGlyArgLeuLeuPheAlaAspProVal 273
 Db 886 TCACAACACTTACGACATGATAAAGCATGATGAGGAGCGCATACCTTA 933
 RESULT 31
 US-10-156-761-413
 / Sequence 413, Application US/10156761
 / Publication No. US20030119018A1
 / GENERAL INFORMATION:
 / APPLICANT: OMURA, SATOSHI
 / APPLICANT: IKEDA, HARUO
 / APPLICANT: ISHIKAWA, JUN
 / APPLICANT: HORIKAWA, HIROSHI
 / APPLICANT: SHIBA, TADAYOSHI
 / APPLICANT: SAKAKI, YOSHIYUKI
 / APPLICANT: HATTORI, MASAHIRA
 / TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 / FILE REFERENCE: 249-262
 / CURRENT APPLICATION NUMBER: US/10/156,761
 / PRIOR FILING DATE: 2002-05-29
 / PRIOR APPLICATION NUMBER: JP 2001-204089
 / PRIOR FILING DATE: 2001-05-30
 / PRIOR APPLICATION NUMBER: JP 2001-272697
 / NUMBER OF SEQ ID NOS: 15109
 / SEQ ID NO 413
 / LENGTH: 5505
 / TYPE: DNA
 / ORGANISM: Streptomyces avermitilis
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (1)..(5505)
 US-10-156-761-413
 Alignment Scores:
 Pred. No.: 0.00907 Length: 5505
 Score: 122.50 Matches: 120
 Percent Similarity: 33.14% Conservative: 50
 Best Local Similarity: 23.39% Mismatches: 174
 Query Match: 6.07% Indels: 170
 Gaps: 20
 DB: 14
 US-09-889-756a-2 (1-412) x US-10-156-761-413 (1-5505)
 Qy 7 LysAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 1083 AGACCGGGAACGCGCGCTGCTCGCTCGCTGAGTCCAAACATCGGTACACGCGAGGC 1142
 Qy 27 LysGly-----GlyAspAlaAlaGln----- 33
 Db 1143 CGCGCGGGGTGTCCCGGTGTGATCAAGATGCTGCTGCGATGACGACGCGGTGTGCC 1202
 Qy 33 ----- 33
 Db 1203 GAAGACCCCTCCACGTGACCGCGCGACCCCGGGGTGACGTGCTGCTGCTGCGTGA 1262
 Qy 34 -----GlyGlyGlnProAlaGlyArgGluAlaProAlaProValGlyVal 49
 Db 1263 CCGTGTACGAGGCGCATGCGGTGCGCGAGACGACCCCTCGCGCGCGCGGGGTCTC 1322
 Qy 50 Val-----ThrValHisProGlnThrValAlaLeuThr--- 60
 Db 1323 GTCTTTCGGATCAGCGCGACCAATGCGCACATATCTCGAGCAGCGCGCCGGAACCGGA 1382
 Qy 61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal--- 79
 Db 1383 GCGCGGGAGACCGCGCCCATGACACGCGTGAAGGCGAGGCGGTGCGCAGGAACTGCTGCG 1442
 Qy 80 -----GlyGlyIleIleGlnLysArgLeuPheGlnGluGlySerTyrValArg 95
 Db 1443 GCGCCGAGCACGAGGTGTGCGGTGATCTCTCCGACGAAGATTGGGCGCGCTGCGAGA 1502

Qy 96 AlaGly-GlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAl 115
 Db 1503 GCAAGCACACGACTGCTG-----TCGTCGGAGCACTCGC 1538
 Qy 115 ArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLeuAlaAspAlaAspLeuAlaAr 135
 Db 1539 GCGCGCTCACCGGCCACGCTCGCGGTCTCGCTGCGCGCACCCGCTCGCGGTGAGCA 1598
 Qy 135 GlyTyrProLeuValAlaAlaGluAlaValAlaSerArgGlnGluTyrAspAlaAlaValTh 155
 Db 1599 CCGGCGCGGTGCTCGCGCCGACGCGG-----GAAGAGCTGCGCGCATCTGAC 1649
 Qy 155 LAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaLysSerAlaGly 175
 Db 1650 CGCC-----CTCGCGCGCGCGCGACCCCGCGCACAGTGTGTACGGGCGCGCCG 1700
 Qy 175 YLeuAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLys 195
 Db 1701 GCGCGACCG-GGCGAAGTGGGTCTCTGCTCGCGGTGAGGTTTC----- 1745
 Qy 195 SValSerGluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGly 215
 Db 1746 -----GAGCGGATCGCATGGGCGCGCATGTTGCA-- 1778
 Qy 215 nThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeuArgAr 235
 Db 1779 -----GGCTACCCCGTTCGCGC 1798
 Qy 235 GglnIleAlaGluGlyLysLeu-----Le 243
 Db 1799 CGGCTTACACGAGGTGTGCGCGCTGTGACTTGGACCTTGGATGTCATGTCGAGCAACGC 1858
 Qy 243 uAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProGly 263
 Db 1859 TTCATCAGACGGGTTCGACGCGACCGCGCTTTCGCGATGAGAGTGGCGCTGTCCGCT 1918
 Qy 263 uLysGlyArgLeuLeuPhe---Ala-AspProValAlaGlnSerThrGlyGlnIle- 281
 Db 1919 TGCCTGAGTGTGTGGCCATCCCGCGGATTAAGTGGCGGTCACTCCGTGGAGAGATTG 1978
 Qy 282 -----ThrLeuAla 284
 Db 1979 CTGCGCGGACGTCGCGCGGGGTGTGTCGCTCGGAGTGGCGGAGCGCTGTGTCGTCG 2038
 Qy 284 rGlnAlaAlaValProAlaAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuM 304
 Db 2039 GTGCGCGGTGATG-----CAGGCGTTCGCCCGCGCGCGCATGTGTGCGGTG--- 2088
 Qy 304 eAspGlnValAlaValAspAsn-----A 312
 Db 2089 -----CAGGCGCGCGAGAGACGAGTGTGCTGCTGCGGATGCGCGGAGCTGCGCGG 2143
 Qy 312 lApheValValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAla 332
 Db 2144 CGATCAACGGCCCTCAAGTGTGTGTGTCGCGCGCCGAGAGCGGGGTATGGCGTT- 2202
 Qy 332 snAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnT 352
 Db 2203 -----GCCGAGGTGTTTCTGCTCAGGCGCGTAAACCTCC 2239
 Qy 352 rIleValThrSer-----GlyLeuLysAspGlyAspL 363
 Db 2240 GGTGAAGGTCACTCAGCGGTTTCATTCCCGCTGATGATGCCAGTGTGGAGAGATTGC 2299
 Qy 363 yValValValGluGluGlyIleSer-----IleAlaGlyIleT 375
 Db 2300 CGGCGGTGTGCGCGGTGTGACGTTGCGTGAAGCGCGGATTCCTGTGTTCTCAACCTCA 2359
 Qy 375 hGlyIly-----AlaLysLysValThrProLysGluThrPalasSerGluAsnGlnAla 393
 Db 2360 CGGCGCGCTGCGCGGCGGTACACCCCGAGACATGGGTCCGCGCATGTCCTCGTGAAGCGG 2419

```

Qy      393 laalaProglInserGlyValGlnThrAlaSerGlu 404
Db      2420 TCCGTTTCGCGATGCGGTGGGACCCCTGCACGAG 2454

RESULT 32
US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Alignment Scores:
Pred. No.:      0.487      Length:      100000
Score:          122.50     Matches:      120
Percent Similarity: 33.14% Conservative: 50
Best Local Similarity: 23.39% Mismatches: 174
Query Match:      6.07%   Indels:      170
                        Gaps:      20
DB:
US-09-889-756a-2 (1-412) x US-10-156-761-15103 (1-100000)

Qy      7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValLeuSerSerGly 26
Db      46019 AGAGCGGCAACGCGCGTGTCTGCTCGCTCAAGTCCAAACATCGGTCAACGACG 46078

Qy      27 LysGly-----GlyAspAlaAlaGln----- 33
Db      46079 CGCGCGGGGTGTCCGCGGTGTGATCAAGATGTGTCTGGCGATGACGACGCGGTCTGCC 46138

Qy      33 ----- 33
Db      46139 GAAGACCTCCACGTCGACCGGCGGCGGTGACGTGTCTGCTCGGTCTGCGTCCGTCGA 46198

Qy      34 -----GlyGlyGlnProAlaGlyArgGlnAlaProAlaProAlaValAlaGlyVal 49
Db      46199 CCTGCTGACGAGGCGATCGGTGTGCGCGAGACGACCACTCCGCGGCGCGGCGGTCTCC 46258

Qy      50 Val-----ThrValHisProGlnThrValAlaLeuThr--- 60
Db      46259 GTCGTTCCGATCAACGCGGACCAATGCGACATCTCTCCGACGACGCGCGCGAACCAGGA 46318

Qy      61 ValGlnLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnVal--- 79
Db      46319 GCGCGGGGACCCGCGCGCATGACACGCGTGAAGCGACGAGGCGCGCGCGCGCAAGCTCTCCG 46378

Qy      80 -----GlyGlyIleIleGlnIleAspLeuPheGlnGlnGlySerIleValArg 95
Db      46379 GAGCGCGACCAACGCGGTGTGCGGTGACTCTCCGCAAGAAATGGGCGCGCGCTGCGAGA 46438

Qy      96 AlaGly-GlnProLeuIleArgIleAspSerSerThrIleGlnAlaLeuGlnSerVal 115
Db      46439 GCAGCGACACGACGATGCTG-----TCGTCCGTCGACCTTCCG 46474

```

```

Qy      115 aArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaIleValAlaAspAlaAspLeuAlaAr 135
Db      46475 GAGCGCGTCACCGCGCCACCGTCGCGTCTGCTGCGCGCCACCGCGCTGCGGTGAGCA 46534

Qy      135 gTyIleProLeuValAlaAlaGlnAlaValSerArgGlnGluIleThrAspAlaAlaValTh 155
Db      46535 CCGGCGCGCGGTCTCTCGCGCGCGCGCGG-----GAAGAGCTGCGCGGATCTTGAC 46585

Qy      155 fAlaIleArgSerAlaGlnAlaGlyValIleValAlaAlaGlnAlaAlaIleValSerAlaG 175
Db      46586 CGCC-----CTCCGCGCGCGCGCGAGCCCGCGCGCACAGGTGTGACGCGCGCGCGCGCG 46636

Qy      175 yIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerIle 195
Db      46637 GAGCGACCG-GAGCAAGTCGGGTCTCTGTCGCGGTCAAGGTTTC----- 46681

Qy      195 eValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgI 215
Db      46682 -----GCAGCGGATCGCGCATGGCGCGCGCGGCGGAGTTGACGA-- 46714

Qy      215 nThrAsnProMetIleValAsnValThrGlnSerAlaSerGluValMetIleValArgAr 235
Db      46715 -----GGCTACCCCGTCTTCCCG 46734

Qy      235 gGlnIleAlaGlnGlyIleValLeu-----Le 243
Db      46735 CGCGCTACACGAGAGTGTGCGCGCTGTGACTTGCCTGCGATGTCATGCGGAAACGC 46794

Qy      243 uAlaAlaAspGlyValIleAlaValAlaGlyIleValPheAspAspGlyThrValIleProGl 263
Db      46795 TTCATCAGACGGGTTCACACGACGCGCGGTGTTCGCGATGAGAGGCGCTGTCCGCT 46854

Qy      263 uIleGlyArgLeuLeuPhe---Ala-AspProValValAsnGluSerThrGlyGlnIle- 281
Db      46855 TGCTGAGTGTGTGCGCATCCGCGGATTAACGTCGCGGATCCTCGTGGAGAAGATTG 46914

Qy      282 -----ThrLeuA 284
Db      46915 CTGCGCGCACAGTGCAGCGGAGTGTGTGCTGCGCGGATGCGCGAGCGAGGTGTGCGGTGC 46974

Qy      284 rGlnAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuIleValArgValLeuM 304
Db      46975 GTGCGCGGTGTGATG-----CAGCGTTCGCCCGCGCGCGCATGTGCGGTG----- 47024

Qy      304 eAspGlnValAlaValAspAsn-----A 312
Db      47025 -----CAGCGCGCGGAGAGACGAGTGTGCTGCGCGGATGCGCGGAGGTGCGCATCCGCG 47079

Qy      312 laPheValValProGlnGlnAlaValThrArgIleValAlaValAspThrValMetIleValA 332
Db      47080 CGATCAACGCGCGCTCAATCGGTGTGCTGCGCGCGCGGAGACGCGGTATGCGGTT- 47138

Qy      332 snAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnT 352
Db      47139 -----GCCAGGTGTGTGCTGCTCAGCGCGCGGAAAGACTTCCC 47175

Qy      352 rIleValThrSer-----GlyLeuIleAspGlyAspL 363
Db      47176 GGTGAAGATCAGTACACCGCTTCATTCGCGGTGATGATGATCCGATGTGAGAGATTCCG 47235

Qy      363 ySValValValGlnGlyIleSer-----IleAlaGlyIleT 375
Db      47236 CGCGGTTTCTCGCGGTCTGACGTTGCGGTGAGCGCGGATTCCTGTGTCTCCAACTCA 47295

Qy      375 hArgIle-----AlaIleValValThrProIleGluIleThrAlaSerSerGluAsnGlnAla 393
Db      47296 CGGCGCGCTGCGCGACCGCTACACCCCGAGCACTGAGTCCGCGACGTCCTGAGAGCGG 47355

Qy      393 laAlaProGlnInserGlyValGlnThrAlaSerGlu 404
Db      47356 TCCGTTTCGCGATGCGGTGCGGACCTTCGACGAG 47390

```

RESULT 33

US-09-308-207-19
Sequence 19, Application US/09308207
Publication No. US20030022323A1
GENERAL INFORMATION:
APPLICANT: MARIA DIAZ-TORRES ET AL.
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT PRODUCTION OF 1,3-PROPANEDIOL
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 4 Cambridge Place
CITY: Rochester
STATE: NY
COUNTRY: U.S.A
ZIP: 14618
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/308,207
FILING DATE: 13-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,601
FILING DATE: 13-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC 369-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-864-7620
TELEFAX: 650-845-6504
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 12145 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PHK28-26
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-308-207-19
Alignment Scores:
Pred. No.: 0.0345 Length: 12145
Score: 121.50 Matches: 80
Percent Similarity: 37.28% Conservative: 49
Best Local Similarity: 23.12% Mismatches: 136
Query Match: 6.02% Indels: 81
DB: 11 Gaps: 17
US-09-889-756a-2 (1-412) x US-09-308-207-19 (1-12145)
QY 69 SerLeuArgThrAlaAspValArgAlaGlnValGlyIleLeuGlnLysArgLeuPhe 88
Db 9662 ACAGTGAATGCCCTTGTCCGGAGTCCGCGAAGTGTATCAGACGCGCATACCTG 9721
QY 89 GlnGluGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyr 108
Db 9722 CGTAAAGAGACTTAAGCGAGAGTCAAGCATGCCGTTAATA----- 9760
QY 109 GlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 9761 ---GCCGGGATTGATATCGCAAC-----GCCACCAAGGAGGCGCGCTGGCG--- 9805
QY 129 AlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGln 148
Db 9806 -----TCCGACTACCGCGACGAGCGAGCGGCTTTGTTCACGCGGATCGTCGCG----- 9853

QY 149 GlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla--- 167
Db 9854 -----ACGACGGGCGATGAAGGAGCGGGAGCAATATCGCGGACCCCTGCGCGCTG 9907
QY 168 ---GlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIle----- 183
Db 9908 GAGCAGGCCCTCGCGAAGAACCCGTGTCATGACGATGCTCTCCGATCTATCTTAAC 9967
QY 184 ---ThrAlaProLieserGlyPheIleGlyGlnSerLysValSerGlyGlyThrLeuLeu 202
Db 9968 GAACCCGCGCGGATGATGCGATGCGCATGAGACCAATCACCGAG--ACCATATTC 10024
QY 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db 10025 ACC-----GAATGACCATGATGCGTCATACCCGAGAG---CCGGCGGGGTGGCC 10075
QY 223 ValThrGlnSerAlaSerGluValMetLysLeuArgArg----- 235
Db 10076 GTT-----GAGCGGGAGACGACTATCGCCCTCGGCGGCTGGCGACGCTGCCGCGGCG 10129
QY 236 GlnIleAlaGlnGlyLysLeuLeuAlaAlaAsp----- 246
Db 10130 CAGTATGCCAGAGGATGATGATCTGATTGACGACGCCCTGATTTCTTGACCGCGTG 10189
QY 247 -----GlyValIleAlaValAlaGlyIleLysPhe 255
Db 10190 TGGTGGCTCAATAGAGCGCTCGACCGGGGATCAAGTGTGGGGGCGGATCTCAAAAAG 10249
QY 256 AspAspGlyThrValTyrProGluLysGlyArgLeuLeuPheAlaAspProValAlaAsn 275
Db 10250 GACACGCGCGTGTGCTGCTGCTG-----AACACCGCGCTCGTAAACCTCGCGGTGTGAT 10303
QY 276 GluSerThrGlnGlnIleThrLeuArgAlaAlaValProAsnAsp-----Gln 291
Db 10304 GAA-----GTAGCGCTGCTGAGCAGGTCCCGAGGGGGTAAATGCGCGCGTG 10351
QY 292 AsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspAlaValAlaValAspAsn 311
Db 10352 GAAGTGGCGCGCGCGCGCGCGAGGTGTCGATCTCTCGAATCCCTACGGGATGCCAC 10411
QY 312 AlaPhe-----ValValProGlnGlnAlaValThrArg 322
Db 10412 TTCTTGGGCTTAAGCCCGGAAGAGACCCAGCGCATCTCCCATCGCGCGCTGATT 10471
QY 323 GlnAlaLysAspThrValMetIleValAsnAlaGlnGlyIleMetGluProAlaGluVal 342
Db 10472 GGCACCGCTTCCGCGGTGCTCAAGACCCCGAGGGGATGTGCAGTGCGCGGTATC 10531
QY 343 ThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAsp 362
Db 10532 CCGCGCGGCG-----AACCTTACATTAGCGGCAAGAGCGCGCGAGAG 10576
QY 363 LysValValValAlaGlnGly 368
Db 10577 GCCGATGTGCGCGAGGCG 10594
RESULT 34
US-10-277-249-1
Sequence 1, Application US/10277249
Publication No. US20030157674A1
GENERAL INFORMATION:
APPLICANT: Emplage, Mark
APPLICANT: Haynie, Sharon
APPLICANT: Lafield, Lisa
APPLICANT: Pucci, Jeff
APPLICANT: Whited, Greg
TITLE OF INVENTION: Improved Process for the Biological Production of 1,3-Propanediol
TITLE OR INVENTION: High Titer
FILE REFERENCE: BC1020 US DIV
CURRENT APPLICATION NUMBER: US/10/277, 249
CURRENT FILING DATE: 2002-10-21
PRIOR APPLICATION NUMBER: 60/149, 534

; PRIOR FILING DATE: 1999-08-08
 ; NUMBER OF SEQ ID NOS: 68
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 1
 ; LENGTH: 12145
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-10-277-249-1

Alignment Scores:

Pred. No.:	0.0345	Length:	12145
Score:	121.50	Matches:	80
Percent Similarity:	37.28%	Conservative:	49
Best Local Similarity:	23.12%	Mismatches:	136
Query Match:	6.02%	Indels:	81
DB:	12	Gaps:	17

US-09-889-756a-2 (1-412) x US-10-277-249-1 (1-12145)

```

Oy 69 SerleuargThrAlaaspValargAlaGlnValGlyGlyIleIleGlnLysargLeuphe 88
Db 9662 ACAGTGATGCCGCTTGTCCGGAGTCGGCGGAAGTGTATCAGACGGCATAGCTG 9721
Oy 89 GlnGlnGlySerTyrValargAlaGlyGlnProleuTyrGlnIleaspSerThrTyr 108
Db 9722 CGTAAAGAGCTAAGCGAGGTCAGATCGCTTAATA----- 9760
Oy 109 GlnLalaenLeuGlnSerAlaargAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 9761 ---GCCGGATTGATATTCGGCAAC-----GCCACCACCGAGGCGCGCTGCG--- 9805
Oy 129 AlaaspAlaaspLeuAlaargTyrLysProleuValAlaAlaGlnAlaValSerargGln 148
Db 9806 -----TCGCACTACCGCGGAGCGGCGGCTTTGTTGCCAGCGGATGTCGCG----- 9853
Oy 149 GluTyrAspAlaAlaValThrAlaLysargSerAlaGlnAlaGlyValLysAlaAla--- 167
Db 9854 -----ACGACGGCGCATGAAGGAGCGGGGCAATATTCGCGGAGACCTCGCGCGCTG 9907
Oy 168 ---GlnAlaAlaIleLysSerAlaGlyLysLeuLeuAlaargSerArgIle----- 183
Db 9908 GAGCAGGCGCTGGCGAACAACCGTGTGATGACGATGCTCTCGCATCTATCTTAAC 9967
Oy 184 ---ThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202
Db 9968 GAAGCGCGCGCGTATGCGGATGCGGATGCGGATGCGAGACCAATCACCAG---ACCTATTC 10024
Oy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleargGlnThrAsnProMetTyrValAsn 222
Db *10025 ACC-----GAATCGACCATGATCGTCATTAACCGCAGACG---CCGGCGCGGCTGAGC 10075
Oy 223 ValThrGlnSerAlaSerGlnValMetLysLeuAlaGly----- 235
Db 10076 GTT-----GGCGTGGGAGCACTATCGCCCTCGGCGGCTGCGCGACGCTGCGCGCGCG 10129
Oy 236 GlnIleAlaGlnGlyLysLeuLeuAlaAlaasp----- 246
Db 10130 CAGTATGCCGAGGGGTGATGCTACTGATACGACGCCCTGATCTTCTTGAGCCGCTG 10189
Oy 247 -----GlyValIleAlaValGlyIleLysPhe 255
Db 10190 TGTGTGCTCATGAGCGCTGACCGGGGATCAACGATGCGGCGGATCTCTCAAAAG 10249
Oy 256 AspaSpGlyThrValTyrProGlnLysGlnLeuLeuPheAlaaspProValValAsn 275
Db 10250 GACGCGCGCGTGTGCTG---AACAAACCGCTCGCTTAACCCCTCGGTGTGATG 10303
Oy 276 GluSerThrGlnIleThrLeuargAlaAlaValProAsnasp-----Gln 291
Db 10304 GAA-----GTACGCTGCTGAGACGAGGCTCCCGAGAGGGGTATGCGCGGCTG 10351
Oy 292 AsnIleLeuMetProGlyLeuTyrValargValLeuMetaspGlnValAlaAspAsn 311
  
```

```

Db 10352 GAATGCGCGCGCGCGGCGAGTGTGCGATCTCTGCAATCCCTACGGATGCCACC 10411
Oy 312 AlaPhe-----ValValProGlnGlnAlaValThrArg 322
Db 10412 TTCTTCGGGCTTAACCCCGAAGAACCCAGGCCATGCTCCCATCGCCCGCCCTGATT 10471
Oy 323 GlyAlaLysaspThrValMetIleValAsnAlaGlnGlyMetGluProArgGlnVal 342
Db 10472 GGCAACGCTTCGGCGTGTGCTCAAGACCCCGAGGGGATGTGACGTGCGGGTGTATC 10531
Oy 343 ThrValAlaGlnGlnGlnGlyThrAsnThrPheValThrSerGlyLeuLysaspGlyasp 362
Db 10532 CCGCGCGGCG-----AACCTTCACTTAAGCGCGGAAAAAGCCGCGAGAG 10576
Oy 363 LysValValValGlnGly 368
Db 10577 GCCGATGTCCCGGAGGCG 10594
  
```

RESULT 35

US-10-213-203-1
 ; Sequence 1, Application US/10213203
 ; Publication No. US20030082756A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BURCH, ROBERT R.
 ; APPLICANT: DORSCH, ROBERT R.
 ; APPLICANT: LAFFEND, LISA ANNE
 ; APPLICANT: NAGARAJAN, VASANTHA
 ; APPLICANT: NAKAMURA, CHARLES
 ; TITLE OF INVENTION: 1,3-PROPANEDIOL AND POLYMER DERIVATIVES FROM A
 ; FILE OF INVENTION: FERMENTABLE CARBON SOURCE.
 ; CURRENT APPLICATION NUMBER: US/10/213,203
 ; PRIOR FILING DATE: 2002-08-05
 ; PRIOR APPLICATION NUMBER: 08/440,293
 ; PRIOR FILING DATE: May 12, 1995
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 1
 ; LENGTH: 12145
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-10-213-203-1

Alignment Scores:

Pred. No.:	0.0345	Length:	12145
Score:	121.50	Matches:	80
Percent Similarity:	37.28%	Conservative:	49
Best Local Similarity:	23.12%	Mismatches:	136
Query Match:	6.02%	Indels:	81
DB:	14	Gaps:	17

US-09-889-756a-2 (1-412) x US-10-213-203-1 (1-12145)

```

Oy 69 SerleuargThrAlaaspValargAlaGlnValGlyGlyIleIleGlnLysargLeuphe 88
Db 9662 ACAGTGATGCCGCTTGTCCGGAGTCGGCGGAAGTGTATCAGACGGCATAGCTG 9721
Oy 89 GlnGlnGlySerTyrValargAlaGlyGlnProleuTyrGlnIleaspSerThrTyr 108
Db 9722 CGTAAAGAGCTAAGCGAGGTCAGATCGCTTAATA----- 9760
Oy 109 GlnLalaenLeuGlnSerAlaargAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 9761 ---GCCGGATTGATATTCGGCAAC-----GCCACCACCGAGGCGCGCTGCG--- 9805
Oy 129 AlaaspAlaaspLeuAlaargTyrLysProleuValAlaAlaGlnAlaValSerargGln 148
Db 9806 -----TCGCACTACCGCGGAGCGGCGGCTTTGTTGCCAGCGGATGTCGCG--- 9853
Oy 149 GluTyrAspAlaAlaValThrAlaLysargSerAlaGlnAlaGlyValLysAlaAla--- 167
Db 9854 -----ACGACGGCGCATGAAGGAGCGGGGCAATATTCGCGGAGACCTCGCGCGCTG 9907
  
```

QY 168 ---GlnAlaAlaIleYsSerAlaGlyIleAsnLeuAsnArgSerArgIle----- 183
 Db 9908 GAGAGAGCCCTGGCGAAACACCGGTGTCGATGACGATGTCCTCCGATCTATCTTAAC 9967
 QY 184 ---ThrAlaProIleSerGlyPheIleGlyIleSerValSerGlyIleThrLeu 202
 Db 9968 GAAGCCGGCGCGGATTTGGGATGGCGATGGAGACCAATACCGAG--ACCATATATC 10024
 QY 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgIleAsnProMetThrValAsn 222
 Db 10025 ACC-----GAATGACCATATCGGTGATACCCGCGAGAG--CCGGGGGGGTGGGC 10075
 QY 223 ValThrGlnSerAlaSerGlyValMetIleuArgArg----- 235
 Db 10076 GTT-----GGCGTGGGACCACTATCGCTCGCGGTGGCGAGCGCTCCGGCGCGC 10129
 QY 236 GlnIleAlaGlyIleValLeuAlaAlaAsp----- 246
 Db 10130 CAGTATGCCGAGGGGTGATGTAATGATACGACCGCTGATTTCTTGACCGCGTG 10189
 QY 247 -----GlyValIleAlaValGlyIleIleSph 255
 Db 10190 TGTGGCTCATGAGCGCTGACCGGGGATCAACGTGGTGGCGGATCTCTCAAAAG 10249
 QY 256 AspAspGlyThrValTyrProGlnuArgIleuLeuPheAlaAspProValValAsn 275
 Db 10250 GACGACCGCGCTGCTGTG-----AACACCGCGCTGGTAACCCCTGCGGTGTGGAT 10303
 QY 276 GluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp-----Gln 291
 Db 10304 GAA-----GTGACGCTCTGAGACAGAGTCCCGAGGGGTAAATGCGCGCGTG 10351
 QY 292 AsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsn 311
 Db 10352 GAATGGCCCGCGCGCGCGAGCTGTGCGATCTCTGCAATCCTTACGGGATGCCAAC 10411
 QY 312 AlaPhe-----ValValProGlnGlnAlaValThrArg 322
 Db 10412 TTCTTCGGGCTAAGCCCGAAGACCCAGGCGATCTGCTCCCGCGCGCGCTGAT 10471
 QY 323 GlyAlaIleAspThrValMetIleValAlaAsnAlaGlnIleGlyMetGluProArgGluVal 342
 Db 10472 GGCACACGTTCCGGGTGTGCTCAAGACCCCGAGGGGATGTCAGTCCGGGTGATC 10531
 QY 343 ThrValAlaGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuIleAspGlyAsp 362
 Db 10532 CCGCGCGGCG-----AACCTCACTATAGCGCGAAGAGCCCGCGAGAG 10576
 QY 363 LysValValValGluGly 368
 Db 10577 GCCGATGTCCCGAGGCG 10594
 RESULT 36
 US-10-193-764-60
 ; Sequence 60, Application US/10193764
 ; Publication No. US20030133943A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
 ; FILE REFERENCE: 1038-1239MIS
 ; CURRENT APPLICATION NUMBER: US/10/193,764
 ; CURRENT FILING DATE: 2002-07-12
 ; PRIOR APPLICATION NUMBER: 09/167,568
 ; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 60
 ; LENGTH: 3543
 ; TYPE: DNA
 ; ORGANISM: Haemophilus influenzae

US-10-193-764-60
 Alignment Scores:
 Pred. No.: 0.00814
 Score: 120.50
 Percent Similarity: 36.05%
 Best Local Similarity: 20.70%
 Query Match: 5.97%
 DB: 12
 Gaps: 18
 US-09-889-756a-2 (1-412) x US-10-193-764-60 (1-3543)
 QY 28 GlyIleAspAlaAlaGlnGlyIleGlnProAlaGlyArgGluAlaProAlaProValVal 47
 Db 2248 GCGCGCATCTCTTGTCTGATGTAATATTTACGCG----- 2283
 QY 48 GlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg--- 66
 Db 2284 AACGCGTTACTGTCTGCAAAATAGCGGTGATTAACCACT--TTGGCAGGCTCTACA 2340
 QY 67 -----LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIle 83
 Db 2341 ATTAAGAACCAACGAGATATACCACTTCAAGTCAATCAGTAAATATCGCGGTAATTT 2400
 QY 84 GlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyIleProLeu----- 100
 Db 2401 TCCGGCAAGCAGTAAC-----GTTAAAGCACTAATAGTTTAACCAACCCAA 2448
 QY 101 -----TyrGlnIleAspSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGln 118
 Db 2449 GCAGACTCAAAATTAATGAGAGCTGAAAGCGAGCTTAATGTAACAGCAAAAGACATA 2508
 QY 119 LeuAlaThrAlaGlnAlaThrLeuAlaIleValAspAlaAspLeuAlaArgTyrLysPro 138
 Db 2509 ATTTGCG-----GATCAATTTCTGTGGGCACTGAAGTAAATCTTACCGGACCGAAGT 2559
 QY 139 LeuValAlaAlaGlnAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysArg 158
 Db 2560 TTAAACACCAACCAAGCTCTAGATTTACTGGAACGAGAGCGTACACTTCAAGCAA 2619
 QY 159 SerAlaGluAlaGly----- 163
 Db 2620 TCAGTAAATATCCGCGCGATGATTTCTGTGGCAAGTAAAGTAAAGTAAAGTAAAGT 2679
 QY 164 ---ValLysAlaAlaGlnAlaAlaIleYsSerAlaGlyIleAsnLeuAsnArgSerArg 182
 Db 2680 TTAAATCTAAATCTCGGTTCAAGATTAAGCAACGCGCGGAGGATGTAACAGT 2739
 QY 183 IleThrAlaProIleSerGlyPheIleGlyIleSerValSerGlyIleThrLeu 202
 Db 2740 GCACAGGTAACATTTGACGTAATTTCCGGTAATACGTAAT-----GTTACAGCA 2793
 QY 203 AsnAlaGlyAspThrThrVal----- 211
 Db 2794 AATATCTGCGATTTAATCTGTTGAAGATGCGCAAAATTTGATCGACAGAGAGCGCG 2853
 QY 212 ThrIleArgGlnThrAsnProMetThrValAsnValThrGlnSerAlaSerGluValMet 231
 Db 2854 ACCCTAAGTCAACGCGGCAAAATTA-----ACCCTAAGCTTGAAGTCTTACAGTACT 2907
 QY 232 LysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaVal 251
 Db 2908 TCAAGTAATTAACCAAGTA-----AACCTTCAAGCTTAAGATGTGACATTTGGCGCA 2958
 QY 252 GlyIleLysPheAspAsp-----GlyThrValTyrProGlnuArgIle 265
 Db 2959 AATATCAATGCTGCTATGTAACATGATTAACAGGCGGCTCTACTACCGGTAAGGCT 3018
 QY 266 ArgLeuLeuPheAlaAspPro-----ValValAsn-----GluSerThr 278
 Db 3019 TCAAGCATTAACGCAACGCGCACCTTGTTATTAACGCAAAAGACCGTGAAGTAAT 3078
 QY 279 GlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAlaIleuMetProGlyLeu 298

Db 3395 CTTGGCGTAAGCGCTGTACGTTTCTGAG 3424

RESULT 38
US-10-156-761-1788
Sequence 1788, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1788
LENGTH: 2088
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2088)
US-10-156-761-1788

Alignment Scores:
Pred. No.: 0.00505 Length: 2088
Score: 119.50 Matches: 97
Percent Similarity: 36.28% Conservative: 63
Best Local Similarity: 22.00% Mismatches: 162
Query Match: 5.92% Indels: 119
Gaps: 20

US-09-889-756a-2 (1-412) x US-10-156-761-1788 (1-2088)

Qy 52 ValHisProGlnThrValAlaLeuThrValAlaLeuPro----- 64
:::|||||
481 CTGCATCCG-----GCCCGCTCGTCGTCAGTCGCGATCGGACCCAGACGGGCTC 534
Qy 65 -----GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGln 78
:::|||||
535 AGGGAGAGTGATGATCTGTGTGCGATGAGTCTGCTCTGCGCGGACGGGCGGACACC 594
Qy 79 Val---GlyGlyIleIleGlnLysArgLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGln 97
:::|||||
595 GTCCAGAGAGGGCCGGTCCGCGACCTTGGCGGACGAGCGGAC-----CGCGCGCGA 648
Qy 98 GlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAla 117
:::|||||
649 CGCGAGCTGAGGAGGAGCGGTGCGGAACCTCCATCCCGTCCGCGCGGAGGAGTTCTGCGCG 708
Qy 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArg----- 135
:::|||||
709 CAGGCGACGCTCTCCGACAGACCTTCACTCCGCGCTGCGCGACCTGACCCGTACCGCGG 768
Qy 136 -----TyrLysProLeuVal 140
:::|||||
769 GACGCGTGTCTGCTCTGCGCTCCGCTACCGCAATCCGCGCATCGGAACCGCTGCTG 828
Qy 141 AlaAlaGlnAlaValSerArgGlnGlnLysAspAlaAlaValThrAlaLysArgSerAla 160
:::|||||
829 GCTGCC---GCCGTGCTTACCTGCTCCCTCGCGCGGTGACGTACCGGCGGTACCGGACCC 885
Qy 161 GlnAlaGlyValLysAlaAlaGlnAlaAlaLysSerAlaGlyLysAsnLeuAsnArg 180
:::|||||
886 GAGGACGCGCGGTGCGGAGACGGGCGCGCCGACCCGCGGCGCGCTTCCGCGCGCTGCTG 945

Qy 181 SerArgIleThrAlaProIleSerGlyPheIleGlnSerLysValSerGlnGlyThr 200
:::|||||
946 TTCAAGTGACCGCGACCGGACGAGCGGCGCTGACGTATGCGTGTACTCCGGAACG 1005
Qy 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
:::|||||
1006 ---ATCGGAAAGAGAGACCGTGTGACACCGCGCGCGCGCGGACGAGACCGATTCGCGC 1062
Qy 221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGlnGly 240
:::|||||
1063 CGGATTCGCTGCTCCAGCGCCGACCGCTACACGAGACTGACCGGCGGCTGCC---GGG 1119
Qy 241 LysLeuLeuAla----- 244
:::|||||
1120 GACATCTGCGCTGCTGCGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1179
Qy 245 -----AlaAspGlyValIleAlaValGlyIle 253
:::|||||
1180 GCGGCTCCCTGCTGCTCGAACCGCGCGACCGCTCCGATCCGCTGCTCCGCGCGCTC 1239
Qy 254 LysPheAspAspGlyThrValTyrProGlnLysGlyArgLeuPheAlaAspProVal 273
:::|||||
1240 GAGCGCGCGACGACGACCC---GACACCGACCGCTTGGCTCGCGCTGCGCGCGC 1290
Qy 274 ValAsnGlnSerThrGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
:::|||||
1291 CTGCGCGAGAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1350
Qy 294 Leu-----MetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsn 311
:::|||||
1351 CTGTCGGGACATGGGCGAGCTGATCTGAGAGGTGGGTGGAGAGATCCGCGCGCGC 1410
Qy 312 AlaPheValAlaPro-----GlnGlnAlaValThrArgGly 323
:::|||||
1411 GGAACGCGCGTCAATGTGCGCGCGCGGAGGTGCGCTTACCGGAGACGCTGCCCGCGCGC 1470
Qy 324 AlaLysAspThrVal----- 328
:::|||||
1471 GTGTCGGGCTCGTGTACCCGCAATGTCAACAGAGACGCGCGCGCGCGCGCGCGCGCGCAT 1530
Qy 329 MetIleValAsnAlaGln-----GlyGlyMetGln 338
:::|||||
1531 GTGCTCTCAGCGAGCGCGCTGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1590
Qy 339 ProArgGlnValThrValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
:::|||||
1591 TTCGCTCGCGCGCTGCTGCGCGCGCGGTGCGCGAGAGTACGTCGCGCGGTGCGCGCGC 1650
Qy 357 GlnLeuLysAspGlyAspLysValValIleGlnGlyIleSerIleAlaGly-----Ile 374
:::|||||
1651 GGCTGCGGGAGC-----GCCCTGCGGAGAGG---CCGCTGCGCGGTACCGCGGTG 1698
Qy 375 ThrGlyAlaLysLysValThr-----ProLysGlnThrAlaSerSer 388
:::|||||
1699 ACCGGCTGAGAGTGTCTGACCGATGAGCGACCCATCCGAAGAC-----TCTCTCC 1752
Qy 389 Gln 389
:::|||||
Db 1753 GAG 1755

RESULT 39
US-10-190-435-51/C
Sequence 51, Application US/10190435
Publication No. US20030143248A1
GENERAL INFORMATION:
APPLICANT: ZUR MEGEDE, Jan
APPLICANT: BARNETT, Susan W.
APPLICANT: LIAN, Ying
APPLICANT: ENGELBRECHT, Susan
APPLICANT: VAN KENSBOURG, Estrelita J.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

FILING DATE: 01-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/453,702
 FILING DATE: 03-DEC-1999
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296.95017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 251-5000
 TELEFAX: (608) 251-9166
 INFORMATION FOR SEQ ID NO: 72:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 46819
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 72:
 US-10-114-170-72

Alignment Scores:
 Pred. No.: 0.362 Length: 46819
 Score: 119.50 Matches: 100
 Percent Similarity: 35.05% Conservative: 50
 Best Local Similarity: 23.36% Mismatches: 165
 Query Match: 5.92% Indels: 114
 Gaps: 20

US-09-889-756a-2 (1-412) x US-10-114-170-72 (1-46819)
 QY 11 AAlaAlaLeuAlaAlaAlaValAlaLeuSerSerCysGlyGlyGlyAsp 30
 Db 32573 GCGCGTGCATATCTGCCAGCGCCGCGAATCGCAACA-----AGTCCGCGGCGAG 32617
 QY 31 AAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaValAlaGlyValVal 50
 Db 32618 GCGTCAAGCAAGTCCACCGCCGCGAATCGCAACAAGTCCGCGATCGTCTGCTTCA 32677
 QY 51 ThrValHisProGln-----ThrValAlaLeuThrValGlnLeuPro 64
 Db 32678 ACAGC-CACACAGAAAGCTGGCGAAGCCACGACAGCGCAGCGAGAGGATTCGCG 32736
 QY 65 GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleGln 84
 Db 32737 TTCGCGAGCGAAGACATCCGA-----AACGACGCGAAGCGTTGGAAAC-----CAG 32784
 QY 85 LysArgLeuPheGlnGlnGlySerThrVal-ArgAlaGlyGlnProLeuThrGlnIleAs 104
 Db 32785 CGCAGAAATCTCAAAAACGCGCTCCCGCATCGTCAGCC-----AG 32823
 QY 104 pSerSerThrThrArgAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAl 124
 Db 32824 TTCGGCGCGCTCATCGCATCTGCGTCTGCTTCAAAAGATGAGCGGACAGACAAAGC 32883
 QY 124 aThrLeuAlaLysAlaAspAlaAspLeuAlaArgThrLysProLeuValAlaAlaGlnAl 144
 Db 32884 GTCAGCAGCGAAGAGCGACGACGCGATCCACGAAAGCGCAGAGCGAGCGCTGGTAG 32943
 QY 144 aValSerArgGlnGlnLysArgAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyVal 164
 Db 32944 TCCGACG-----GCGACGCTCAAGAGCAAAAGTACGGCGGAATCTGCAGC 32988
 QY 164 LlysAlaAlaGlnAlaAlaIleLysSerAla-----GlyIleAsnLeuAs 179
 Db 32989 AACCGCGCTGACAGACGCGAAGCGGACAGAGATTTGCAATCCGCGCGGCGCTTGA 33048
 QY 179 nArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnG 199
 Db 33049 GGAATCGAGCAGCGAAGAA-----AAGGGATAGTACAGCTCAGAGTGCAGCTAACAG 33102

QY 199 YThr-----LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAs 217
 Db 33103 CACTTCGAGACACTGGCGGAAGCGCAAAAGCCGTTAAGGCCCGATGATGAGCTGCTAA 33162
 QY 217 nProMetCysVal-----As 222
 Db 33163 CGGGAATATACCCGACAGAGATGACAGACAGACAGAAAGGATGATTCAGCTTAAGCA 33222
 QY 222 nValThrGlnSerAlaSerGlnValMet-----LysLeuArgAr 235
 Db 33223 CGGCACCAACAGCAGCATCTGAATGCTGGCGGCAAGCCAAAGTCGTTAAGGCGACCTTA 33282
 QY 235 gGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPh 255
 Db 33283 TGACCTTGCTAACGGGAATATATCTGCTCAGGACGCTACAGACAGCA----- 33328
 QY 255 eAspAspGlyThrValLysProGlnLysGlyArgLeuLeuPheAlaAspProValValAs 275
 Db 33329 -----CAAAAGGAATGTCAGCTCAGTGTGCAACCAACAG 33366
 QY 275 nGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeu 295
 Db 33367 CGCATCTGAA-----ACGCTT-----GCCGACACCGAAA----- 33397
 QY 295 tProGlyLeuThrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValVa 315
 Db 33398 -----GCAGTGAAGCAGCTATGATATATG----- 33424
 QY 315 lProGlnAlaValThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnG 335
 Db 33425 -----AATGTCGGGTACCTTCTGCCCGTAAGGTGAATGTGAAGGC 33465
 QY 335 Y-----GlyMetGlnProArgGlnVal-----ThrVa 344
 Db 33466 GCTTTCATCGGATATTAACCTGACGCGCAAGATATGTAAGCTTAATCAACACAGAT 33525
 QY 344 lAlaGlnGlnGlnGlyThrAsnTrp-----IleValThrSerGlyLeuLysAspGlyAs 362
 Db 33526 GTCAATCAGCGGTGCTGCTGTTGTTCAAAATTAAGCAAGGTAACTGCCAGCGGAG 33585
 QY 362 pLysValValValGlnGlyLysSerIleAlaGlyIleThrGlyAlaLysLysValThrPr 382
 Db 33586 TTCTGTGTT-----TCAATTACGTGATTGTGCGCGGATTTAACGTGGGTCAAC 33639
 QY 382 cLysGlnTrpAlaSerSerGln 389
 Db 33640 TCAACAGCAGGTATATCTGAA 33661

Search completed: September 8, 2003, 15:11:57
 Job time : 23898 secs

THIS PAGE BLANK (USPTO)


```

Qy 48 G1yValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeu 67
Db 163 GGTGTAATGTTGCTCAACACCAAGAGTGTGAACAAACCGTTGACGCTTTCAGCCGTAAT 222
Qy 68 GluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeu 87
Db 223 TCACCAATATCAAAATTTCTGAAGTTCCTCAACCAAGTGGCGGATTTTAAACGGTTA 282
Qy 88 PheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThr 107
Db 283 TTTCTGTAAGGAAGACTGTTGTCGTAAGGTCAGCGCTTTATGAGCTTCGACTTGAAGC 342
Qy 108 TyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAla 127
Db 343 AACCGTCGACGTAAGAAATATGCAAAAGCATCTCTCAACAACAAGCAAAATCTAGCT 402
Qy 128 LysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAlaValSerArg 147
Db 403 TCACCTACGTAACCAAGTAAATCGTTATTAACACCTGTTCTAGTAATGCTGTGCTAAA 462
Qy 148 GlnGlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla 167
Db 463 CAGGAATATGATGACTTATCTTGCTCAAGTCAATGTTGCAAGCAACAAGTTCAGACGCT 522
Qy 168 GlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIle 187
Db 523 AAGCTCAAGTCAACAAATGCAAAATGATGATCTTGTTATTTCAAAATTCGCTCTCAT 582
Qy 188 SerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThr 207
Db 583 TCTGCGCAATCGTGTCTTCTTCAAGTACGCTGCTGTTGCTTACTGCAAAACGACGCT 642
Qy 208 ThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAla 227
Db 643 GACCGCTTGTAAGCATTCACCAAGTATGATCTTATGATATTAATCAAGTCAAT 702
Qy 228 SerGlnValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGly 247
Db 703 GCTAAGTATATGCTTACGTAACAACCACTAAGGCAAGTAAATTAACAGTAACAC 762
Qy 248 ValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProGlnLysGlyArgLeu 267
Db 763 ACG---AAAGTAAATTAAGCTTGAAGATGTTCTCAATCTCAATCGAAGGCGCAAT 819
Qy 268 LeuPheAlaAspProValValAsnGlnSerArgGlnIleThrLeuArgAlaAlaVal 287
Db 820 GCTTCTCTGACGCTTCTGTAACCAAGATACAGAACATTAATTAACGTCGATATC 879
Qy 288 ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnVal 307
Db 880 TCTAACCAGATCAATTTATGCTTCCGGTATGTAACCACTGGCGCAAAATGTTCCAGGC 939
Qy 308 AlaValAspAsnAlaPheValValProGlnGlnAlaValThrArg-----GlyAlaLys 325
Db 940 GTTGTTCCAAAATGCTTACCTGATTCCTCAAGCTGCAATTAATCTGTTAACCAGACAA 999
Qy 326 AspThrValMetIleValAsnAlaGlnGlyLysGlnProArgGlnValThrValAla 345
Db 1000 GCTGTAGGATGCTGTTATATGCTAAGGGGTTGAGACCGTCTGTTGAAACCTCT 1059
Qy 346 GlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLeuVal 365
Db 1060 GGTGTTCAGAGCAAACTGATTTGTGACTAAGCGCTTAAAGCGCGGATTAAGTCAAT 1119
Qy 366 ValGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrProLysGlnTrp 385
Db 1120 GTTATATGCT-----GTTGCCAAAGTTAAAGAGGCGCAAGATATCAGCAAAACCTTAT 1173
Qy 386 AlaSer-----SerGlnAsnGlnAlaAlaIlePro-----GlnSer 397
Db 1174 CAAGCTCAACAGCAAACTCTCAAGGTCGACAGCAACAAATGCTCGCAAAACCGGCTCAATCA 1233
Qy 398 G1yValGlnThrAlaSerGlnAlaLysThrAlaSerGlnAla 411

```

```

Db 1234 GGT---AAACCTCAAGCAGAAACAGAAAGCAAGCTTCAATCA 1272
RESULT 2
US-09-252-991A-444/C
; Sequence 444, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 444
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-444
Alignment Scores:
Pred. No.: 1,15e-72 Length: 1251
Score: 753.00 Matches: 174
Percent Similarity: 58.60% Conservative: 68
Best Local Similarity: 42.13% Mismatches: 125
Query Match: 37.30% Indels: 46
Gaps: 10
US-09-889-756a-2 (1-412) x US-09-252-991A-444 (1-1251)
Qy 8 AlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu---ValLeuSerSerCysGly 26
Db 1137 GCCATGCGTATGACTGTTCCGCGCTGCTGCGGATTTCCGCGCTTCCGGGTGCGGA 1078
Qy 27 LysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaProVal 46
Db 1077 AAAAGC-----GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1054
Qy 47 -----ValGlyValValThrValHisProGlnThrValAlaLeuThrValGlu 62
Db 1053 CAAGCGCGGAGGTCGGGATCGTACCCCTGGAAGCGAGACGGTGAACCTGAATACCGAG 994
Qy 63 LeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIle 82
Db 993 CTGCGCGGCGCGGACCAATGCGTTCGCGATCGCGAGCGTGGTCCCGCAGTGAAACGGCATC 934
Qy 83 IleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGln 102
Db 933 ATCTCAAGCGCTGTTCAAGAGGAGGAGCAAGCTCAAGCGCGGCGAGCTTACAG 874
Qy 103 IleAspSerSerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAla 122
Db 873 ATGACCCCGCCACCTTCAGAGCGGACTACCAAGAGCCCGCAAGCACTGCTGCGCAC 814
Qy 123 GlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAla 142
Db 813 CAG-----GAGCAGCGCCAGCGCTACAAAGCTGCTGCTGCGCGAC 775
Qy 143 GlnAlaValSerArgGlnGlnTyrAspAlaAlaValThrAlaLysArgSerAlaGluAla 162
Db 774 CAGCGCTGAGCAAGCAGCGATC-----GCCAGCGCCAAATGCC 736
Qy 163 G1yValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 735 GCTTACTGACATCAAGCGCGGCGGTGAGACGCGCGGATCAACCTGCGCTACACCAAG 676
Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202

```

```

Db      675 GTGCTGTCGGCATCTCCGGCCGATCGGCGCTTCGCGGTGACCGAAGCGCCCTGCTG 616
Qy      203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgInThrAsnProMetTyrValAsn 222
Db      615 ACCAAGCGGCGAGCGCAAGCGATGCCGCGCAACAGCTCGAACCCGATCTACGTCGAC 556
Qy      223 ValThrGlnSerAlaSerGluValMetIleuArgArgGlnIleAlaGluGlyLysLeu 242
Db      555 GTCAACCGAGCGCTCACCGCGCTGCTGCGCTGCGCGCGAAGTGGCGAGCGGCGCAAGTTG 496
Qy      243 ---LeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyr 261
Db      495 GAGCGCGCGCGCGCAACGCGCGAAGGCTCTCCCTGAAGCTGAGAGAGAGGTAAGCAATAC 436
Qy      262 ProGluLysGlyArgLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db      435 CCGCTGGAAGGTCCGCTCGAATTTCTCCAGGTTTCGTCGACGAAGGACCGGCTCGGTC 376
Qy      282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db      375 ACCATCGCGCGCGCTTCCCAACCGCAACAGAGCTGTCGCGGCGATGTTGTTTAC 316
Qy      302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
Db      315 GCGCAGTTGCAAGAGCGCTCAACGCAAGGCAATCTCTGCTCGCGACGAAGCGGTGACC 256
Qy      322 ArgGlyAlaLysAsp-----ThrValMetIleValAsnAlaGlnGlyMetGluPro 339
Db      255 CGCAGCTCAAGAGCGCGAGCTACCGCGCTGCTGTGAACGCGCGAAGCAAGGTGACGTCG 196
Qy      340 ArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrIleValThrSerGlyLeuLys 359
Db      195 CGGGATTCAGAGCGCGAGCGGCTGATCGCGCAAAATGCTGTTACCGAAGGCTGTAAC 136
Qy      360 AspGlyAspLysValValValGluGlyIleSerIleAla-----GlyIleThrGlyAla 377
Db      135 GCGCGCGCAAGATCATTCAGAAAGCGCTGCGATTCGTGACCGCGGCTGTCAG---GTG 79
Qy      378 LysLeuValThrProLysGluTyrPalasSerSerGlnAsnGlnAlaIleAlaProGlnSer 397
Db      78 AAGACCGTCCGCGGCAAGATGTCGCTCCGCGCAAGAAAGCGCGCGCTCGG----- 25
Qy      398 GlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db      24 -----GCGAAAACGACGACGCAAG 7

RESULT 3
US-09-252-991A-417
; Sequence 417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 31142
; SEQ ID NO 417
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-417

Alignment Scores:
Pred. No.: 1,32e-72 Length: 1362
Score: 753.00 Matches: 174
Percent Similarity: 58.60% Conservative: 68
Best Local Similarity: 42.13% Mismatches: 125

```

```

Query Match: 37.30% Indels: 46
DB: 4 Gaps: 10
US-09-889-756a-2 (1-412) x US-09-252-991A-417 (1-1362)

Qy      8 AlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu---ValLeuSerSerCysGly 26
Db      226 GCCATGCGTGTACTGTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 285
Qy      27 LysGlyLysAspAlaAlaGlnGlnGlyGlnProAlaIleArgValAlaProAlaProVal 46
Db      286 AAAAGC-----GAGGCGCGCGCGCGCG 309
Qy      47 -----ValGlyValAlaThrValHisProGlnThrValAlaLeuThrValGlu 62
Db      310 CAAGCGCGGAGGTCCGCGATGTCGATCCCTGGAAGCGAGACCGGTGACCTGAAATACGAG 369
Qy      63 LeuProGlyArgLeuGlnSerLeuArgThrAlaAspValAlaArgAlaGlnValGlyIle 82
Db      370 CTGCGCGCGCGCGCAACATGCGTTCGCGATCGCGAGGTGCGTCCCGAGTGAACGCGATC 429
Qy      83 IleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGln 102
Db      430 ATCTTCAAGCGCTGTTCAGAGAGGCGAGCGACGTCAGAGCGGCGAGAGCTTACCGAG 489
Qy      103 IleAspSerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnIleLeuAlaThrAla 122
Db      490 ATGACCGCGCGCGCACCTACGAGCGCGACCTACGAGCGCGCGCAACCTGCTTCGAC 549
Qy      123 GlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrIleProLeuValAlaAla 142
Db      550 CAG-----GAGCAGCGCGCGCGCTACCAAGCTGCTGCGCGAC 588
Qy      143 GlnAlaValSerArgGlnGlnIleTyrAspAlaAlaValThrAlaLysAspSerAlaGlnAla 162
Db      589 CAGCGCGTGAAGAGCAGCGATC-----GCCGACCGCAATGCC 627
Qy      163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db      628 GCGTACTCGTAGCTTCAAGAGCGCGGTGAGCAGCGCGCGATCACTGCTACACCAAG 687
Qy      183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202
Db      688 GTGCTGCGCGCATCTCCGCGCGCTGCGCGCTGCGCGGTGACCGAAGCGCGCTGCTG 747
Qy      203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db      748 ACCAAGCGGAGCGCAACGCGATGCGCACCGCTGCAACAGCTCGACCGATCTACGTCGAC 807
Qy      223 ValThrGlnSerAlaSerGluValMetLysLeuArgGlnIleAlaGluGlyLysLeu 242
Db      808 GTACACCGAGCGCTCCACCGCGCTGCGCTGCGCTGCGCGCGACTGCGCGAGCGCGCACTG 867
Qy      243 ---LeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyr 261
Db      868 GAGCGCGCGCGCGCAACGCGCGAGGCTCTCCCTGAAGCTGAGAGCGGTGACCAATAC 927
Qy      262 ProGluLysGlyArgLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db      928 CCGCTGGAAGGTCCGCTCGAATTTCTCCAGGTTTCGTCGACGAAGGACCGGCTCGGTC 987
Qy      282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db      988 ACCATCGCGCGCGCTTCCCAACCGCAACAGAGCTGCTGCGCGCATGTTGTTTAC 1047
Qy      302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr 321
Db      1048 GCGCAGTTGCAAGAGCGCTCAACGCAAGGCAATCTCTGCTCGCGACGAAGCGGTGACC 1107
Qy      322 ArgGlyAlaLysAsp-----ThrValMetIleValAsnAlaGlnGlyMetGluPro 339
Db      1108 CGCAGCTCAAGGCGCGAGCTACCGCGCTGCTGTGTGAACGCGCGCAAGAGGTGACGTCG 1167

```

Qy 340 ArglnuValThrValAlaGlnGlnGlnGlnThrAsnTrpIleValThrSerGlyLeuIys 359
Db 1168 CGGGATCAAGGCGCGGCGGTGATCGGCAACAATGCTGTTACCGAAGGCGCTGAC 1227
Qy 360 AspGlyAspIysValIleValGluGlyIleSerIleAla-----GlyIleThrGlyAla 377
Db 1228 GCCCGCGCAAGATCATTCACGAAGGCGCTGCGAGCGCGGTGTCAG-----GTG 1284
Qy 378 LysIysValThrProIysGluTrpAlaSerSerGluAsnGlnAlaIleAlaProGlnSer 397
Db 1285 AAGACCGGCGCGGCAAGATGTCGCTCGCGCAAGAGCCGACGCCGCTCCG----- 1338
Qy 398 GlyValGlnThrAlaSerGluAlaIysThrAlaSerGlu 410
Db 1339 -----GCCGAAACCGACACGCAAG 1356

RESULT 4
US-09-453-702B-102/c
Sequence 102, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-Dec-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 7886
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-453-702B-102

Alignment Scores:
Pred. No.: 8.87e-66 Length: 7886
Score: 701.50 Matches: 163
Percent Similarity: 60.10% Conservative: 66
Best Local Similarity: 42.78% Mismatches: 135
Query Match: 34.74% Indels: 17
DB: 4 Gaps: 7
US-09-889-756a-2 (1-412) x US-09-453-702B-102 (1-7886)

Qy 7 LysAlaMetArgAlaIleAlaIleValAlaIleValAlaLeuSerSerCyGly 26
Db 6823 AAAATATATGAAGTATATACCAATCTGTAGTGCATGCTGCTTATCTGTGTG--- 6767
Qy 27 LysGlyGlyAspAlaIleGlnGlyGlnProAlaGlyIleValGluAlaProAlaProVal 46
Db 6766 -----GATATATACCAAGTATACATTTCCCGTCAGAAACAGAA----- 6725
Qy 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
Db 6724 GTCGGCGTGTATAGTCAATTCACACCGGTCTCGGTATGCTGTAATTAACCGACGC 6665
Qy 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnIysArg 86
Db 6664 ACCAGTCTCGCTCACTGCGCAAGTACGTCCGAGGTTGGGGAAATTATCCAGAAACGC 6605
Qy 87 LeuPheGlnGluGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSer 106
Db 6604 TTATTTAAGAGAGGTGATCTGCTCAAGCTGACAGCCGCTTACAGATTGATGCGGCC 6545
Qy 107 ThrTyrGlnAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaIleThrLeu 126
Db 6544 AGTATCAGCTGCATGCAATGAAGCTCGGCGACGATTAACAACAGCAGGCA---CTG 6488
Qy 127 AlaIysAlaAspAlaAspLeuAla---ArgTyrIleProLeuValAlaIleAlaVal 145
Db 6487 GTAAACCCGATTCGCCGAAGAGCGCACGCTTATCCCGCATGTGGAAAGAGAACGGTGT 6428
Qy 146 SerArgGlnIleTyrAspAlaIleValThrAlaIysSerAlaGluAlaGlyValIys 165
Db 6427 TCACACAGAGATGCTGATGATGCTCAGTCACTGTCACAGATTAAGCCAGAGTGGCG 6368
Qy 166 AlaAlaGlnAlaIleIleYsSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
Db 6367 GCGAAAGAGCCGACGGAAGTCTGCGCATTAATCTTACGAGACCAAGGTAACCGCA 6308
Qy 186 ProIleSerGlyPheIleGlyGlnSerIysValSerGluIleThrLeuLeuAsnAlaGly 205
Db 6307 CCGATTTCAGGGCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6248
Qy 206 AspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGln 225
Db 6247 CAAGATACAGGTTAAGCACTATTCGTGATGATGATGATGATGATGATGATGATGAT 6188
Qy 226 SerAlaSerGluValMetIysLeuArgArgGlnIleIleGluGlyIleLeuAlaIle 245
Db 6187 TCCAGTGTGATTTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6140
Qy 246 AspGlyValIleAlaValGlyIleIysPheAspAspGlyThrValTyrProGluIysGly 265
Db 6139 AGTACACCATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6080
Qy 266 ArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAla 285
Db 6079 GGTCTGGAATCAACCGAAGTCTGCTGATGATGATGATGATGATGATGATGATGATGAT 6020
Qy 286 AlaValProAsnArgGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAsp 305
Db 6019 ATTTTCCCAATCCACACAGCAATTAATACCGGGAATGTTGTCGCTGCTGCTGCTG 5960
Qy 306 GlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaIys 325
Db 5959 GAAGCGCTGATGAAGAGCGCTATTTCTGCGCGCACAGGCGCTGACGCGGATCTTAA 5900
Qy 326 Asp-----ThrValMetIleValAsnAlaGlnGlyIleMetGluProArgGluValThr 343
Db 5899 GGCATGCAACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5840
Qy 344 ValAlaGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuIysAspGlyAspIys 363
Db 5839 ACGGAGAAACATATGCTGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5780

```

Qy 364 ValValValGluGlyLeuSerIleAlaGlyIleThrGlyAlaValValValThrProLys 383
Db 5779 CTGATGTTGAAGT-----TCGCGCAAGTCACTTCAGGAGCGATCAAGCTGTT 5726
Qy 384 Glu 384
Db 5725 GAA 5723

RESULT 5
US-09-252-991A-14019
; Sequence 14019, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14019
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14019

Alignment Scores:
Pred. No.: 1e-66 Length: 1233
Score: 698.50 Matches: 164
Percent Similarity: 62.66% Conservative: 81
Best Local Similarity: 41.94% Mismatches: 130
Query Match: 34.60% Indels: 17
Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-14019 (1-1233)
Qy 10 ArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyIleGlyGly 29
Db 94 AGGATCGGGCGGTGGCTATGGCCATCGC-----TTGGCGGGTGTGGGCC-GGGCGA 146
Qy 30 AspAlaIaGlnGlyGlnProAlaGlyArg-GluAlaProAlaProAlaValGlyVa 49
Db 147 AGAGCGCA-----GGAGCGCGCGCAATGCTGTGGC--GTGAGGT 188
Qy 49 ValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGln 69
Db 189 CTGACGCGTCAGGCGCGCCGCTGAGTTCGAACTGCTGGCGGATCGAAC 248
Qy 69 rLeuAlaGlyThrAlaAspValAlaGlnValGlyIleIleGlnIleArgLeuPheG1 89
Db 249 GGTGGCTGTGCGGAGGTGCGCGCGCGGTGGCGCGGATGCTGTGGCAAGCGCTTCA 308
Qy 89 ngInGlySerTyValArgAlaGlyGlnProLeuTyArgGlnIleAspSerSerThrTyrg1 109
Db 309 GAGAGGGCGCGAGCTCAAGCTGCGCACTGCTGTTCAATGATCGGACCGCTGA 368
Qy 109 uAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysa1 129
Db 369 GGTGGCGGTGTCCGCGCGCGAGGTGAGCTGCGCGCGCAACCGCGGTGCTGTTCAGGC 428
Qy 129 aAspAlaAspLeuAlaArgTyArgProLeuValAlaAlaGlnAlaValSerArgGlnG1 149
Db 429 GAGGCGCGGTGCTGCTGCTACAGCGCGTGTGAAGATCCAGCGCGTCAAGCAGCA 488
Qy 149 uTyAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAl 169
Db 489 CTTCGATACCGCACCGCGCACTGCGCGAGCGCGCGCAACCGCGGTGCGCAAGG 548
Qy 169 aAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerG1 189

```

```

Db 549 CGACCTGGAACCGCGCGCTGAACCTCGGCTACGCTCGGTCACTCGCGCATCTCCGG 608
Qy 189 yPheIleGlyIleSerTyValSerGlyIleThrLeuAlaAlaGlyAspThrThrVa 209
Db 609 GCGCATCGCGCGCGCTGTGTACCGAGCGCGCGCTGTGTGGCGAGCGCGAGCGCT 668
Qy 129 uAlaAlaThrIleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerG1 229
Db 669 GATGGCGCGATCAGAGATTCGATCGATCTATGGGATTTCAACCAAGCCGCGCGCA 728
Qy 229 uValMetLysLeuAlaArgGlnIleAlaGlyIleLysLeuLeuAlaAlaAspGlyVal11 249
Db 729 GGGCTGCGCGCTCGCGACCGCTTGAAGAAAGACCTTGGCCGCGCGCACAGCCAG-- 786
Qy 249 eAlaValGlyIleLysPheAspArgIleThrValTyProGlnLysGlyArgLeuPhe 269
Db 787 -GGCTGACCTTCGGGTC--GAGGAGCGCTTACAGACCGCGCGCGCTTGCAGTT 842
Qy 269 eAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAs 289
Db 843 CGCGACGTGCGCGGTGATCGCGTACCGGCGAGATCCCTTGGCGCGCAAGTCCGCA 902
Qy 289 nAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaVa 309
Db 903 CCGCGACGGGCTCTGCTCGCGCGCATGACGTGCGCATGACGCCCGCGCGCATGCA 962
Qy 309 laAspAlaAlaPheValValProGlnIleAlaValThrArgGlyAlaLysAspThr----- 327
Db 963 CAACGAGCGATCTGTGTCGCGCGCAAGCGCGCGTGCACCGCTCCAGCGAGCGCGCCA 1022
Qy 328 -ValMetIleValAlaAlaGlnGlyIleMetGluProArgGluValThrValAlaGlnG1 347
Db 1023 GGTGATGTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082
Qy 347 ngInGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValG1 367
Db 1083 GCAGGGTTCGCGCTGCGCATCAACGAGGCGCTGAGCGCGGTGACCGGCTATGTCGG 1142
Qy 367 uGlyIleSerIleAlaGlyIleThrGlyAlaLysValThrProLys---GluTrpAl 386
Db 1143 CGGC-----CTGGCTGCGTGCAGCGCGCGGTGAAGATCTGCGCAAGCGGATGTC 1196
Qy 386 aSerSerGluAsnGlnAlaAlaAlaProGln 396
Db 1197 CCAGCGCGCAACCGCATCTGCGCGCGCA 1227

RESULT 6
US-09-252-991A-14117/c
; Sequence 14117, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14117
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14117

Alignment Scores:
Pred. No.: 4.5e-65 Length: 1161
Score: 683.00 Matches: 159
Percent Similarity: 62.60% Conservative: 77

```



```

QY 194 SerIysValSerGIuGIYThrLeuLeuAsnIaGIaAspThrThrValIleuAlaThrIle 213
Db 697 TCTTTTGTCACTGACATGTGTCGTGCGTCAAGCGCATACCAATACCATGCGCAACCATT 756
QY 214 ArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGIuValMetIysLeu 233
Db 757 CAACAGATTGATTAAGTCTATGTTGATGATTAAGCAATCATGTTAGTACGTATGACGGCTTA 816
QY 234 ArgArgGlnIleAlaGIuGIYIysLeuLeuAlaIaAspGIYValIleAlaValGIYIle 253
Db 817 CAAGTCGGCTACAAAGCGGTGA---TTATCAGCAATAGTACAAACCGTTCGATT 873
QY 254 LysPheAspAspGIYThrValTyrProGIuIysGIYAspLeuPheAlaAspProVal 273
Db 874 ACCAATAGCCAGCGCAACCTTAATACGTCAACAGCAAAATGTTGTTGAAGATTATTAAT 933
QY 274 ValAsnGIuSerThrGIYGIuIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 934 GTTACCCCGGAACACGGCGATGTCACATTCGTATTGAAATTAACTACGACGAAAGCAAAA 993
QY 294 LeuMetProGIYLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 994 TTACTTCGGCGCATGATGTCGTGTCATATGATGATGCTGCTTCTATTCTTCAAGCGCTA 1053
QY 314 ValValProGIuGlnAlaValThrArg----GlyAlaIysAspThrValMetIleVal 331
Db 1054 TTGGTTCGGCGCAAGCGATCCACGTAATATCATGCGCGACCTCAGTATATGATCATC 1113
QY 332 AsnAlaGIuGIYGIuMetGIuProArgGIuValThrValAlaGIuGIuGIYThrAsn 351
Db 1114 AATCCCAAGGTACAGCGGAATTCGTCTTATGCAAAATGGACAGCAATATGACCACTTC 1173
QY 352 TrpIleValThrSerGIYLeuIysAspGIYAspIysValValIleGIuGIYIle 369
Db 1174 TATATTGCTAACAAAGGCTTGAAGTCGTGATCAAAAGTCGTGTGTAAGGATT 1227

```

RESULT 8

```

US-09-252-991A-11404
; Sequence 11404, Application US/09252991A
; Patent NO. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11404
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11404

```

Alignment Scores:

```

Pred. No: 1.51e-60 Length: 1245
Score: 642.00 Matches: 157
Percent Similarity: 55.36% Conservative: 65
Best Local Similarity: 39.15% Mismatches: 157
Query Match: 31.80% Indels: 22
DB: 4 Gaps: 5

```

US-09-889-756A-2 (1-412) x US-09-252-991A-11404 (1-1245)

```

QY 14 LeuAlaIaIaValAlaValIleuSerSerGIYGIYAspAlaIaGln 33
Db 91 CTGGCGGCGCTGTGCGCTTCTCTGCGGTGCGAAGACAGCGGAC----- 141
QY 34 GlyGIuIleProIaGIYArgGIuAlaProIaIaProVal---ValGIYValIleThrVal 52

```

```

Db 142 -----GCCGGAAGACTGCGAGAGCCCCCGCCAGGTCGCGTATCTGCC 189
QY 53 HisProGlnThrValAlaIleuThrValGIuIleProGIYArgLeuIleuSerLeuArgThr 72
Db 190 AGCGCGGCGCTTTCGGCATCACCGAGCTGCGGAGCGCTGGAAGGATACCGCAG 249
QY 73 AlaAspValArgAlaGlnValGIYIleIleGlnIysArgLeuPheGlnGIYSer 92
Db 250 GCTGAATGGGGGGCGCGCTGCGCGGATGTCACCCCTGCTGTACGAGGAAGCGCAG 309
QY 93 TyrValArgAlaGIYIleuProIleuTyrGlnIleAspSerSerThrTyrGIuAlaAsnLeu 112
Db 310 GACGTCCGCGCGCGACCGGCTGTTCCAGATGACCTTCGCGCTTGAAGCGGCGCTG 369
QY 113 GluSerAlaArgAlaGlnIleuAlaThrAlaGlnAlaThrLeuAlaIysAlaAspAlaAsp 132
Db 370 GACATCACTGCGCGCGCTGCGCGCGCGAGCGCAACCGCGCGCGCGCAG 429
QY 133 LeuAlaArgTyrIysProLeuValAlaAlaGlnAlaValSerArgGlnGIYIysAspAla 152
Db 430 CTGAAGCGCTACCGCGCATGATCAAGACCGCGCATACGCAACCGCATACCGAA 489
QY 153 AlaValThrAlaIysArgSerAlaGIuAlaGIYValIysAlaAlaGlnAlaIleIys 172
Db 490 GCCCAGACCGACCGCGCGCGCGCGCGCGCGCGCATGCTCGCGCAAGCGCAACTGAG 549
QY 173 SerIaGIYIleAsnIleuAsnArgSerArgIleThrAlaProIleSerGIYIleGIY 192
Db 550 CAGCGCGCGCTGCGCGCTGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
QY 193 GlnSerIysValSerGIuGIYThrLeuLeuAsnAlaGIYAspThrThrValIleuAlaThr 212
Db 610 CCGCGCGCTGTACCGAAGCGCGCTGTGCGGAGAGACTCGCGCACCTCTGACCGCG 669
QY 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGIuValMetIys 232
Db 670 GTGACACAGATCGATCGATCTACGTGAACCTTCTCCACCGCGCGCGAGTCGCGCG 729
QY 233 LeuArgArgGlnIleAlaGIuGIYIysLeuLeu---AlaAlaAspGIYAlaIleAlaVal 251
Db 730 ATCCACGGCGCGATTCGCGCAAGCGCGAGAGGTGTGCGCGCAAGACATGCGCGT 789
QY 252 GlyIleIysPheAspAspGIYThrValTyrProGIuIysGIYArgLeuPheAlaAsp 271
Db 790 CGCTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 849
QY 272 ProValValAsnGIuSerThrGIYGIuIleThrLeuArgAlaAlaValProAsnAspGln 291
Db 850 CTGGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 909
QY 292 AsnIleLeuMetProGIYLeuTyrValArgValLeuMetAspGlnValAlaValAspAsn 311
Db 910 CGGGAATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
QY 312 AlaPheValValProGlnIleAlaValThrArgGIYAlaIysAspThrValMet---Ile 330
Db 970 GCGATCACCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029
QY 331 ValAsnAlaGlnGIYGIYMetGIuProArgGIuValThrValAlaGlnGlnGIYThr 350
Db 1030 GTCAACCCACAGGGCTGTGTGAAGACGTGAGGTCCGCGCGCAACCGTGCAGGCGCG 1089
QY 351 AsnTrpIleValThrSerGIYLeuIysAspGIYAspIysValValIleGIYIleSer 370
Db 1090 GACTGATCATCAGCAGCGCGCTCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1149
QY 371 IleAlaGIYIleThrGIYAlaIysIysValThrProIysGIYTrpAlaSerSerGIuAsn 390
Db 1150 -----CAGCATGCGCGCGCGCTCAGCGT 1173
QY 391 GlnAlaIaIaIaProGlnSerGIYValGlnThrAlaSerGIuAlaIysThrAlaSerGIu 410

```


Db 1174 AAGCGGATGATCCGCGACGCGGCGACGCGCGCCCTCAGCCGCTGCGCGC 1233

QY 411 Ala 411
|||

Db 1234 GCG 1236

RESULT 9

US-09-252-991A-11488/c
; Sequence 11488, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11488
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11488

Alignment Scores:

Pred. No.: 4.04e-58 Length: 1104
Score: 619.00 Matches: 145
Percent Similarity: 57.64% Conservative: 70
Best Local Similarity: 38.87% Mismatches: 148
Query Match: 30.66% Indels: 10
DB: 4 Gaps: 4

US-09-889-756A-2 (1-412) x US-09-252-991A-11488 (1-1104)

QY 20 LeuValLeuSerSerCysGlyLyseGlyValAspAlaIaIaGlnGlyGlnProAlaGly 39
Db 1104 CTATTCTCGTGGGCTGGAGAGACGCGAC-----GCCGGG 1066

QY 40 ArgGluAlaProAlaProVal---ValGlyValValThrValHisProGlnThrValAla 58
Db 1065 AAGCTGGGAGGCGCCCGCGAGGTCGCGCTGATCTGCGCAGCGCGCGCTATCGGC 1006

QY 59 LeuThrValGluLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGln 78
Db 1005 ATCAACACGCGAGCTGCGCGGACGCTGGAACGCTGACCGCCAGCGCTGAAGTGGGCGCGC 946

QY 79 ValGlyGlyLeuLeuGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGln 98
Db 945 GTGCGCGGCATCTGACCCCTGCTGCTTACGAGAGAGCCAGAGACGTCGCGCGCGCAC 886

QY 99 ProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGln 118
Db 885 GTGCTGTCTCCAGATCGACCTGCGCCCTTGAAGCGCGCTCGACATCGACGCGCGCGC 826

QY 119 LeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLysPro 138
Db 825 GTGCGAC 766

QY 139 LeuValAlaIaIaGlnAlaValSerArgGlnGlnTyrAspAlaIaIaValThrAlaLysArg 158
Db 765 CTGATCAAGACCGCGCGCATCAGCGAGCGAGCGATCACCCGAAAGCGACGACGCGCGC 706

QY 159 SerAlaGluAlaGlyValLysAlaAlaGlnAlaIaIaLysSerAlaGlyLysAsnLeu 178
Db 705 CAGGCGCTGCGCGAGATGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 646

QY 179 AsnArgSerAlaGlyLeuThrAlaProLysSerGlyPheIleGlyGlnSerLysValSerGlu 198
Db 645 GGCTACGCGACGATCACCGCGCGCGATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGAA 586

[illegible]


```

QY 75 ValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlySerThrVal 94
Db 2 GTGCGGGGGCGCGTGGCGCGATCGTCAACCGCTGTCAGGAGAGCGCAGAGCGTC 61
QY 95 ArgAlaGlyGlnProLeuThrGlnIleAspSerSerThrTyArgLysAlaLeuGlnSer 114
Db 62 CGCCCGCGGACCGGCTGCTTCAGATCGACCTCGCCCTTGAAGGCGCGCTCGACATC 121
QY 115 AlaArgAlaGlnLeuAlaThrAlaGlnAlaLeuAlaLysAlaAspAlaAspLeuAla 134
Db 122 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181
QY 135 ArgTyArgProLeuValAlaAlaGlnAlaValSerArgGlnGlyTyArgAlaAlaVal 154
Db 182 CGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 241
QY 155 ThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAla 174
Db 242 ACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 301
QY 175 GlyLeuAlaLeuAlaArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSer 194
Db 302 CGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361
QY 195 LysValSerGlyGlyThrLeuLeuAlaGlyAspThrThrValLeuAlaThrIleArg 214
Db 362 CTGTGTCACGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 421
QY 215 GlnThrAspProMetTyArgAlaAsnValThrGlnSerAlaSerGlyValMetLysLeuArg 234
Db 422 CAGATCATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT 481
QY 235 ArgGlnIleAlaGlyLysLeuLeu--AlaAlaAspGlyValIleAlaValGlyIle 253
Db 482 CGGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 541
QY 254 LysPheAspArgGlyThrValTyArgProGlyLysArgLeuPheAlaAspProVal 273
Db 542 GTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601
QY 274 ValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 602 GTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
QY 294 LeuMetProGlyLeuThrValArgValLeuMetAspGlnValAlaValAspAlaPhe 313
Db 662 TTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 721
QY 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAspThrValMet--IleValAsn 332
Db 722 ACCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
QY 333 AlaGlnGlyLysMetGlyProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTyr 352
Db 782 CCACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841
QY 353 IleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIleAla 372
Db 842 ATATCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
QY 373 GlyIleThrGlyAlaLysValThrProLysGlyLysProLysSerGlyLysAlaGlnAla 392
Db 896 -----CAGATGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 925
QY 393 AlaAlaProGlnSerGlyValGlnThrAlaSerGlnAlaLysThrAlaSerGlyAla 411
Db 926 GTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 982

```

RESULT 11
 US-09-328-352-444
 ; Sequence 444, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:

```

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAIUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 444
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-444

Alignment Scores:
Pred. No.: 9,32e-40 Length: 1230
Score: 451.50 Matches: 131
Percent Similarity: 51.45% Conservative: 82
Best Local Similarity: 31.64% Mismatches: 180
Query Match: 22.36% Indels: 21
DB: Gaps: 7

US-09-889-756a-2 (1-412) x US-09-328-352-444 (1-1230)
QY 4 TyrAlaPheLysAlaMetArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSer 23
Db 31 TTTCACATGCTGCTCCATCTTGTGCTATTGTAGCAACCGGTGGAGATTATGTTG--- 87
QY 24 SerCyGlyLysGlyGlyAlaAlaAlaGlnGlyGlnProAlaGlyArgGlyAlaPro 43
Db 88 ---TTACATGAAATGCGCGATGCGAAAGTGCACCAACCGCTCCCAACGATGCTACT 144
QY 44 AlaProValAlaGlyValValThrValIleProGlnThrValAlaLeuThrValGlnLeu 63
Db 145 GTTGATGTACCCCGCAGTAGTACG-----AAACCACTTACCGATTGGCGAAGATAT 195
QY 64 ProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIle 83
Db 196 TCGGTCGTTTAAAGCAATGATCAAGTTGATTTGGCTCAAGTTTCAGGAAACTT 255
QY 84 GlnLysArgLeuPheGlnGlySerTyArgAlaGlyLysGlnProLeuTyArgIle 103
Db 256 ATTCGCGTACATTTCAGAGTGAAGCGCTCGTAAAAAGGTGATTTACTTTTCACATC 315
QY 104 AspSerSerThrTyArgLysAlaLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGln 123
Db 316 GACCTCGCTCTTTAAGCAGACTGAACCGTGCAGAAAGCCCACTTGTCTAGCTGAA 375
QY 124 AlaThrLeuAlaLysAlaAspAlaAspAlaArgTyArgProLeuValAlaAlaGln 143
Db 376 GCACAGGTAAATATACCGCAAGCATCTTGGCGTATTCAGTCTCATTCAGAGTAT 435
QY 144 AlaValSerArgGlnGlyTyArgAlaAlaValThrAlaLysArgSerAlaGlnAlaGly 163
Db 436 GCTGTTCCTCCGCAACATCGATTTAGCCGAATATGATGACGTTTCAGGAAATGCTTAC 495
QY 164 ValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIle 183
Db 496 CTACAGACCGCTAGACTGCTGCTCATCTGACGTTTAATTTAATAATACACCGGTAT 555
QY 184 ThrAlaProIleSerGlyPheIleGlyLysSerValSerGlyLysThrLeuLeuAsn 203
Db 556 ACAGACATGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 615
QY 204 AlaGlyAsp--ThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyArgAla 222
Db 616 GCAGGTAAAGCGCGCAAGGTTTAAAGTTAGTGTGTATACGCGCTATACGATCT 675
QY 223 ValThrGlnSerAlaSerGlnValMetLysLeu-----ArgArgGlnIleAlaGln 239
Db 676 TTTCGATGTGATGAACAACTTACCTGAATATATATCATGATATGAGGTATTCAGACAA 735
QY 240 GlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThr 259

```


QY 406 LysThrAlaSerGluAlaGlu 412
 Db 165 AAGGTGGCGGCTCAAGGAC 145

RESULT 13

US-09-252-991A-15065
 ; Sequence 15065, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 15065
 ; LENGTH: 1368
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-15065

Alignment Scores:

Pred. No.:	2,666-39	Length:	1368
Score:	448.00	Matches:	133
Percent Similarity:	51.11%	Conservative:	75
Best Local Similarity:	32.68%	Mismatches:	177
Query Match:	22.19%	Indels:	22
		Gaps:	9

US-09-889-756a-2 (1-412) x US-09-252-991A-15065 (1-1368)

QY 14 LeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLysGlyLysAlaAlaAlaGln 33
 Db 172 CTCGCACTCGCGCCGCTGCTGCTGAGCGCTGCGCAAGGCCCGCAACACCCCA 211
 QY 34 GlyGlyGlnProAlaGlyArgGluAlaProAlaProValAlaGlyValAlaThrValHis 53
 Db 232 GGC-----ATGGCGCGCCCGCAAGGTCAGCGCTCGCGCAAGTCATC 270
 QY 54 ProGlnThrValAlaLeuThrValGlnLeuProGlyValArgLeuGlnSerLeuArgThrAla 73
 Db 271 GAACAAACCCCTGAACGAGTGAAGCAATTACCGCGCTGAGAGGCCCGGAGTCGCTG 330
 QY 74 AspValArgAlaGlnValGlyGlyLeuLeuGlnLysArgLeuPheGlnGlnGlySerTyr 93
 Db 331 GAGCTGGCGCGCGGGGTGGGTGCTGATGACCGCGCTGCTTCATGAAGCGGACATG 390
 QY 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGln 113
 Db 391 GTGAAGAAAGCCACCTGCTGCTTCAGATGACCGCGCCGCTTCAGAGCGGAGTCAAG 450
 QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeu 133
 Db 451 CGCCTGAACCCGAGGAGCAACGAGCGCGCGCCCGCAAGCGCGCTCAACGAGCC 510
 QY 134 AlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGlnTyrAspAlaAla 153
 Db 511 CACGGCGGCAAGCGCTGCGCGCGCAAGCGGATCTCCCGGGAATTCCCGCGCGCCCG 570
 QY 154 ValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173
 Db 571 ACCACCGCGCGCCGAGGAGCAAGCGCGCTGCGCGCGCAAGCGCGCAACTGAGCGCG 630
 QY 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln 193
 Db 631 GCGCGCTGAACCTTGAAGCTTCAACCGGATCAACCGCGCGGATCAACGCTCGCGTCAAGCCG 690
 QY 194 SerLysValSerGlnGlyThrLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213

Db 691 GCCAGATCACCCCGCAACCTGTCACCTCCGGGAA--ACCTGCTCACCACTCTG 747
 QY 214 ArgGlnThrAspProMetTyrValAlaValThrGlnSerAlaSerGluValMetLys--- 232
 Db 748 GTACGACACGACAGGTCTACGCTTACTTCGACCGCAAGAACCGCTGCTCCCAAGTAC 807
 QY 233 -----LeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
 Db 808 CTCGACTGGCCCGCCAG-----GCCGTGCGGACAGCGGACGAGCCCGGCTTAC 861
 QY 251 ValGlyIleLeuPheAspAspGlyThrValTyrProGlnLysGlyArgLeuLeuPheAla 270
 Db 862 CTCGCGCTGAGCAGCAGGACGCGAAC-----CCGACCTGGCGCGCTGACTTCTTC 915
 QY 271 AspProValValAsnLysSerThrGlyGlnIleThrLeuArgAlaAlaAlaProAsnAsp 290
 Db 916 GACACACAGTCAACCCCGGCTACCGGACCATTCGCGCGCGCTGTTGCAACAGCC 975
 QY 291 GlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAsp 310
 Db 976 AAGGGGAGTTTACCCCGGGGCTCAACGTGCGCTGAAGTTGTCGCGCAAGACTTAC 1035
 QY 311 AsnAlaPheValValProGlnGlnAlaVal--ThrArgGlyAlaLysAspThrValMet 329
 Db 1036 GCGCCACCTGATCAAGAGCAAGGCGTGGACCGACTGGGCAAGAGTTGCTGCTG 1095
 QY 330 IleValAsnAlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlnGly 349
 Db 1096 GTCTGGAAGGGGACCAACAGACCTTACCGGACCGTGAAGATGGACCGAAGCTGGAG 1155
 QY 350 ThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIle 369
 Db 1156 GGCTGCGGATCGTCCGACGCGGCTGAGCAAGGGCGAGCCGATCTCTGATGATGCTG 1215
 QY 370 SerIleAlaGlyIleThrGlyAlaLysLysValThrProLys-----GlnTrpAlaSer 387
 Db 1216 CAG-----CGGGTCCGCGCGGCGATGAGTGCATCCGAGAAAGTGCAGATGCGCAG 1269
 QY 388 SerGluAsnGlnAla-----AlaAlaProGlnSerGlyValGlnThrAlaSerGluAla 405
 Db 1270 GCCGACACCTTGGCCACCTCTGCGCGCTGCGGACGTGCTGCGCAGACGCAACCAACCG 1329
 QY 406 LysThrAlaSerGluAlaGlu 412
 Db 1330 AAGGTGGCGGCTCAAGGAC 1350

RESULT 14
 US-09-252-991A-14539
 ; Sequence 14539, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 14539
 ; LENGTH: 1245
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-14539

Alignment Scores:
 Pred. No.: 2,726-37 Length: 1245
 Score: 429.00 Matches: 119
 Percent Similarity: 46.35% Conservative: 59

Best Local Similarity:	30.99%	Mismatches:	172
Query Match:	21.25%	Indels:	34
DB:	4	Gaps:	7

US-09-889-756A-2 (1-412) X US-09-252-991A-14539 (1-1245)

Oy	6	theuysalameArg-----Alaalalaleuhalalalaleu	20
Db	8	TTCCGTACCCTCGCCCGCTGATCACC GCCCTTGCCGCCCGTATCGGCTG	67
Oy	21	Valleu-----SerSerCySgIySgIyAAspalaalagIngly	34
Db	68	GTGATGTGGCTGGCGCGCCGCTCGACCGTCTCCGACGGGGACCCGGTCCGGG	127
Oy	35	GIyGIInProAlaGIyArg-----GIyAlaProAlaProValGIyAlValThraI	52
Db	128	GGCAAGCCGGGCGCGCGCTGCCAAGGCCAAGCGCTCACCGTGGCGGTGGCAAGGTrG	187
Oy	53	HisProGIInThraIaleuThraIValGulaeProGIyArgIeugIuSerIeuaArgTrH	72
Db	188	GAGCAGGGCGACCTGGCGCTGCATTTCAACGGCGTTGGACCGCTCACCGCTTCAACAGC	247
Oy	73	AlaAspValArgAlaGIInValGIyGIleIeGIInIySArgIeubheGIInGIySer	92
Db	248	GTGAACGTCAACCGCGGGGTCAACGGCGAGCTGTGTCAAGTGTGTTCAGAGAGGGCAG	307
Oy	93	TyrValAArgAlaGIyGIInProIeuTyrGIInIleAspSerSerThyTrGIuAlaenIeu	112
Db	308	GAGGTCAAGCGCGGACCTGTGGGGGTGTGCACCGCGCATCAAGAGGGCGCTG	367
Oy	113	GIuSerIleAArgAlaGIInIeuAlaThraIaGIInAlaThreIuAlaIySAlaAspAlaAsp	132
Db	368	GCCCAAGCGCGAGGGCACGCTGTATGCAGAACAGAGCCCACTGAAGAAGCCGAGATGCAC	427
Oy	133	IeuAlaArgTyrTyrIySProIeuValAlaAlaGIuAlaValSerArgGIInIyTrAspAla	152
Db	428	CTGCAGCGCTCAAGGGGCTGTACCGCGGAGCTCATGTCCCAAGCAGACCCCTGTGATACC	487
Oy	153	AlaValThraIalalyArgSerAlaGIuAlaGIyValIyAlaAlaGIInAlaIleIyS	172
Db	488	CAGGAAGCCCGAGGTCCGCGATTGCCAGGGCACCATCTGTACCAACAGGGCCAGGTTCGAC	547
Oy	173	SerAlaGIyIleAsnIeuAsnArgSerArgIleThraIaProIleSerGIyPheIleGIy	192
Db	548	GACGCGCGCTCAATCTGACCTTACCCAGAGTCCGGGACCCATTTCCGGGGCGCTCGGC	607
Oy	193	GIuSerIySValSerGIyGIyThreIuIeuAsnAlaGIyAspThrThraIleuAlaTrn	212
Db	608	CTGCAGCAGGTGGACATCCGGAACCTGTGCACCAAGGGGAGATCCACGCGCGTGTGGTGG	667
Oy	213	IleArgGIInThraIAspProMetTyrValaenValThGIuSerAlaSerGIuValMetIyS	232
Db	668	ATCACCACAGTGAAGCCGATCTCGGTGTGTTCAGCTGCGCGACGACAGACAGATCGGCACC	727
Oy	233	IeuArgArgGIInIle---AlaGIuGIyIySleuIeuAlaIaIa-----	245
Db	728	GTGTCTGAGACAGATGAAGACGCCCGCGCAAGCTGGCGGTACCGCGTCAAGCCGCAACAG	787
Oy	246	AspGIyValIleAlaValGIyIleIySPhleAspAspGIyThraIyTrProGIuIySgIy	265
Db	788	GACCAAGGTTCTCGCC-----GAAGGC	808
Oy	266	ArgIeIeuPheAlaAspProValValaenGIuSerThrGIyGIInIleThreIuArgAla	285
Db	809	ACCCTGACCAACCTGTGACACCAAGATGCACACCAACCGGACCGGTCAAGCTCAAGGCG	868
Oy	286	AlaValProAsnAspGIInAsnIleIeuMetProGIyIeuTyrValArgValIeueMetAsp	305
Db	869	CGCTTCGAGAAAGCGGACGCGAAGCTGTTCGCCCAACCAAGTTGCTCAACGTGCGCCGTGCTG	928
Oy	306	GIuValAlaIalAspAsnAlaPheValValProGIInAlaIalValThraArgGIyAlaIyS	325
Db	929	GCGCATACCTCAAGGGCGGTGCTGACCATTTCCAGCAAGCGCGTGTGAGCGGGCAACAC	988

Qy	326	AspIhr---	-ValMetLeIeVlaAAlagInglYclYmetGUpProArGrGlUvalThrVal	344
				:::
Db	989	GGrATCTATgTgTAAGCGGTGGCGCCGAACAAGGTGCACCGAGCGAGCGTCATC		1048
Qy	345	AlaIngInglngInglYThrAntPrlleValThrsErGlyLeuLysAspGlYAspLysVal		364
		:::	:::	:::
Db	1049	GGCACCAGCAAGAACGAGCGGGTGtGTGTGAAAGCGGCCCTGAAGGCCGCGCACAGCATGTG		1108
Qy	365	ValValGluglY		368
Db	1109	GTGGTGGAGAAGC		1120

RESULT 15
110 00 353 0018 14670

; Sequence 14670, Application US/09252991A

; GENERAL INFORMATION:

```

; APPLICATION: MATC O. RUDEMIERU ET AL
; TITLE OF INVENTION: NUCLEIC ACID ANALYSIS

```

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196-136

CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR APPLICATION NUMBER: US 60/094,190

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 14670
;
; LENGTH: 1290

```

```

;
; TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

```

US-09-252-991A-14670

Alignment Scores: .

Score:	429.00	Match
--------	--------	-------

Best Local Similarity: 30.99%

Query Match:	21.25%	100%
DB:	4	Gaps

US-09-889-756A-2 (1-412) X US-09-252-991A-1

Our 6 Photos

Alignment Scores:

Fixed. NO.:	42.00e-37	Length:	1230
Score:	429.00	Matches:	119

Percent Similarity:	46.35%	Conservative:	59
Best Local Similarity:	30.99%	Mismatches:	172

Query Match:	21.25%	Indels:	34
DB:	4	Gaps:	7

[illegible]

1000

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

US-09-889-756A-2 (1-412) x US-09-252-991A-14670 (1-1290)

Oy		6	PheIysalamecArg-----AlaIalaIeuaIaIaIaValaIaleu	20
			:::	
Db		37	TTCGATACCTGGCCGCCGTGATCAACCGCCTTGCGCTTGC CGCGCGCGTATCGCGCTG	96
			:::	
Oy		21	Valleu-----SerSerysgIylsgIylAspAlaIagIngly	34
			:::	
Db		97	GTCATGTGGCTGGCGGGCCCGCCTTCGACCCGTCCTCCAGCGGGCACC CGCGTGCgggC	156
			:::	
Oy		35	GlylgnProAlaglAryg-----GlualaProAlaprovalValglYvalThrVal	52
			:::	
Db		157	GGCAAGCGCGCGCGCGCGCTGCCAAGGCCAACGGCTCACGTCGGCGGTGCCAGGGTG	216
			:::	
Oy		53	HispProGlnThrValAlaLeuthValgluleuProglArygleuGluSerleuArythr	72
			::: :::	
Db		217	GAGCAGGGCGACCTGGCGCTGCATTCCAAACCGCGCTTGCCACCGCTTCAAACCG	276
			:::	
Oy		73	AlaaBpyAlarAglaIagInvalGlYglYlletleGlnLysArgleuPheGlnGluYser	92
			::: :::	
Db		277	GTCMAAGTCMAAGCCCGGGGTCAACGGCAGCTGTCTCAAGAGTCTTTCTTCAGAGGGGCAG	336
			:::	
Oy		93	TyrValaIrgalaglYlgnProleuTyrgInleLysSerSerThrTYrgInualaIsnleu	112
			:::	
Db		337	GAGGTCAAAGCGCGCGACTCTGTGGGTGGTGCACCGCGCACTCAAGAGCGCGCGCTG	396
			:::	
Oy		113	GluSerAlaArAglaIagInleuAlaThrAlaGlnAlaThrleuAlaIyAlaAspAlasp	132
			:::	
Db		397	GCCCCAGGCGCGGCGACGCTGATGCAGAACCAGCGCACTGAAGAACCGCGAGTTCAC	456
			:::	

Db 408 CGCTTCGAGAAACCGCAGCGAGCTGTTCACCAACGATTCGTCACGTCGCTGCTG 349
Oy 306 GlnValAlaValAspAsnAlaPheValAlaProGlnGlnAlaValThrArgGlyAlaLys 325
Db 348 GCGCATACCTCCAGAGCGGTGCTGACCATTCACGCAACGCCGTGACAGCGGACCAAC 289
Oy 326 AspThr---ValMetIleValAsnAlaGlnGlyMetGluProArgGluValThrVal 344
Db 288 GGTATCTATGTGTACGTGTGCGGCGCCGACCAACAAAGTCAGCAGCGAGCGTGCATC 229
Oy 345 AlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyVal 364
Db 228 GGCACACAGCAGAACGACGCGGTGTGTGAAAGCGCTGAAAGCGCGGACGAGCTG 169
Oy 365 ValValGlnGly 368
Db 168 GTGGTGAAGGC 157
RESULT 17
US-09-252-991A-7520/c
Sequence 7520, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7520
LENGTH: 1386
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7520
Alignment Scores:
Pred. No.: 1,14e-33 Length: 1386
Score: 396.50 Matches: 126
Percent Similarity: 47.55% Conservative: 58
Best Local Similarity: 32.55% Mismatches: 171
Query Match: 19.64% Indels: 32
DB: 4 Gaps: 7
US-09-889-756a-2 (1-412) x US-09-252-991A-7520 (1-1386)
Oy 28 GlnGlyAspAlaValGlnGlyGlnProAlaGluArgGluAlaProAlaProVal--- 46
Db 1236 GCGCGTGGGAGATGCTGCTGCGGCGGACCGCGGAGGAGCGCGCGCGCGCGCGCGG 1177
Oy 47 -----ValGlyValAlaThrValHisProGlnThrAlaAlaLeuThrValGlnLeuPro 64
Db 1176 GTCCCGGTGAACGTGGCGCGGTGAGAGCGGCGACGTCGACGAACAGTCAGGCGCATC 1117
Oy 65 GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
Db 1116 GGCACGCGGACCTTGTTCACACAGCTGTATCCGACCCAGATCGACGCGAGTTGACC 1057
Oy 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnGlnProLeuTyrGlnIleAsp 104
Db 1056 GCGCTGCTGTGAGCGAAGGCGAATGTGTGAGCGGCGGCGAGTTGCTGCGACATTCAT 997
Oy 105 SerSerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 996 GACCGCGCGCTGTGCTGCGCGCTGAGAGAGCTCAGGCTCCAGGCGAGCAACCGAGCC 937
Oy 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAla 144
Db 936 CAGCTGAATTCGCGGAGCAAGAGACCTGCAACGCTACCGACGCTGTATGCGGACGTCGCG 877

Oy 145 ValSerArgGlnGlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyVal 164
Db 876 GTTCGCGCCCAACGCTGCGACACGACGACGACGACGACGACGACGACGACGACGAC 817
Oy 165 LysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThr 184
Db 816 AAGGCCAAGATGCGACCATTCACAGCGAGCGGCGCTGTCTTACACCCGATCACC 757
Oy 185 AlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeuLeuAla 204
Db 756 TCGCGGTATTCGCGAAGTCGTGTATCCGCAACCTTCATGTGCAACCTGTGCGGTC 697
Oy 205 GlyAspThrThrValIleAlaThrIleArgGlnThrAsnProMetTyrValAsnValThr 224
Db 696 GCGCACAGTCTCGGCGCTGTTCAGGTACGACCAAGTCGACGATCTCGATCTTCTCC 637
Oy 225 GlnSerAlaSerGluValMetLysLeuArgArgGlnIle-----AlaGlnGlyLys 241
Db 636 CTGCACAGCAAGACGTTGCCCGACGTTGACAGCGCTCGCGCGGCGAGCGCGCTGCGC 577
Oy 242 LeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgIleThrValTyr 261
Db 576 GCTTACAGCGCGACGCGGAGCGCGCTG3GC----- 544
Oy 262 ProGluArgGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIle 281
Db 543 -----GAGGCGCGGTGCTGACCATTCACACCAACGATTCACGATTCACGATTC 490
Oy 282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301
Db 489 CGCGTGGCGCTCTTCGATTAACCGCGAGCGCGCTGCGCGCGGCGAGTCTGTCGCG 430
Oy 302 ValLeuMetAspGlnValAlaValAspAsnAlaPheValAlaProGlnGlnAlaValThr 321
Db 429 GTGAGCTTGCAACCGCGGTCGCGCGGACGACTGTGCTGTCGACCAAGCGGTGCGC 370
Oy 322 ArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyMetGluProArgGlu 341
Db 369 GCGCGCTGAGAGGCAATTCGTCATCCGGGTGCGCGCGACCGCGGTGAGCGGTCGCA 310
Oy 342 ValThrValAlaGlnGln---GlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
Db 309 GTGCGGCGTCTCCAGCATTCGACGAGCGCTCAGT-----GTGGTGAAGCGCTGCGCAGC 256
Oy 361 GlyAspLysValValAlaGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysVal 380
Db 255 GGTACCAAGGTGTGTGACGCGGCACTCG-----CGGCTG 220
Oy 381 -ThrProLysGluTrpAlaSerSerGluAsnGlnAlaAla-----ProGlnSer 397
Db 219 ATGCGCGCGCGCTGTGTCATTCAGAGACCGCGCGGACCTGTGCTCAGCGCGGAG 160
Oy 397 GlnValGlnThrAlaSer 403
Db 159 CCGCAGCGGTGACCGCGCG 141
RESULT 18
US-09-252-991A-7817
Sequence 7817, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 7817
LENGTH: 1509
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7817

Alignment Scores:

Pred. No.:	1,48e-33	Length:	1509
Score:	396.00	Matches:	123
Percent Similarity:	47.47%	Conservative:	55
Best Local Similarity:	32.80%	Mismatches:	168
Query Match:	19.61%	Indels:	29
	4	Gaps:	6

US-09-889-756a-2 (1-412) x US-09-252-991A-7817 (1-1509)

```
QY 28 GIVGLYASPLAIAAGLNGIVGLINPROIALAIVRGLUWALAPROVAL--- 46
DB 421 GGCAGTTGGGATGCTCTCGCGCGACGCCCGGGCGAAGCGCGCGCGCGCGCAGG 480
QY 47 -----VALGIVVALThrValHisProGlnThrValAlaLeuThrValGluLeuPro 64
DB 481 GTCCCGGTGAACGTGGCGCGGTGAGCGCGCCACGTCACGACCAACAGTCAGCGCATC 540
QY 65 GIVARGLIeugIuSerIeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
DB 541 GGCAGGTGACTTCTGTCACCAACGTGTGTATCCGACCCGAGTCGACGGCGCATGACC 600
QY 85 LVARGLIeugIuSerIeuArgThrValArgAlaGlnValGlyIleIleAsp 104
DB 601 CGCTGTGTGTGACCGAAGGCGAGATGTGAGCGCGCGCATGTCGCGGACCATCATGAT 660
QY 105 SerserThrIuAlaAsnIeuGluSerAlaArgAlaGlnIleuAlaThrAlaGlnAla 124
DB 661 GACCGCCCGCTCGTCGCGCGGTGAGACAGCTCAGGCTTCAGGCGGAGCAACAGCGCC 720
QY 125 ThrIeuAlaIysAlaAspAlaAspLeuAlaArgIuIysProIeuValAlaAlaGlnAla 144
DB 721 CAGCTGAATTCGCGCGACGACGACTCAACGCTACCGACCTGTATGCGGAGCGTGGC 780
QY 145 VALSerArgIuGlnIuIrrAspAlaAlaValAlaThrAlaIysArgSerAlaGlnAlaGlyVal 164
DB 781 GTGTGCGCGCAACTGTCGACCGACGACGACGCGCGCATGTCGACGCGGACCGCTG 840
QY 165 LVAlaIaIaGlnAlaAlaIleIysSerAlaGlyIleAsnIeuAsnArgSerArgIleThr 184
DB 841 AAGGCCAACATGCTACCATCAACGCGCGGCGTGCCTGTCTACACCGCGATCAC 900
QY 185 ALAProIleSerGlyPheIleGlyGlnSerIysValSerGluGlyThrIleuLeuAsnAla 204
DB 901 TCGCCGGTATCCGCGCAAGTCGGTATCCGCAACGTGATCGGCAACCTGTGCGGTC 960
QY 205 GLYAspThrThrValIeuAlaThrIleArgGlnThrAsnProMetIrrValAsnValThr 224
DB 961 GGGCAGACGTCCGCGCTTCAGCGTACGCCAGATCCGACGATCTCCGATCTTCTCC 1020
QY 225 GlnSerAlaSerGluValMetIysIeuArgArgGlnIle-----AlaGluGlyIys 241
DB 1021 CTCGACACGAGAACGTTGCCCGCTGTGACAGCGCGCTGTCGCGCGACGCGCGGTGCGC 1080
QY 242 LeuIeuAlaIaAspGlyValIleAlaValGlyIleIysPheAspAspGlyThrValTyr 261
DB 1081 GCTTACAGCGCGCGCGCGCGCGCGCGCTGCGC----- 1113
QY 262 ProGluIysGlyArgLeuIeuPheAlaAspProValValAsnGluSerThrGlyGlnIle 281
DB 1114 -----GAGGCGCGGTGTCGACCATCGAACCAATCGACATTCACCGGTATCACATC 1167
QY 282 ThrIeuArgAlaIaValProAsnAspGlnAsnIleIeuMetProGlyIleuIrrValArg 301
DB 1168 CGGTGCGCGCTCTTCGATACCGCGACGCGCGGTGCGCGCGGATTCGTCGCGC 1227
QY 302 ValIeuMetAspGlnValAlaValaIaAspAsnAlaPheValValProGlnGlnAlaValThr 321
```

```
DB 1228 GTGAGCTTCGACACCGGGGTCCGCGCGACCAAGTTGGTCTGTTCGACCAAGCGGTGCGC 1287
QY 322 ARGIVAlaIysAspThrValMetIleValAlaAlaGlnIleGlyMetGluProArgIu 341
DB 1288 CGCGCGCTGAGGACCAATTTCTGTCACCGGGTCCGCGCGACCGCGCGCGGTGCCA 1347
QY 342 VALThrValAlaGlnGln---GlnGlyThrAsnThrIleValThrSerGlyLeuIysAsp 360
DB 1348 GTCCGGGTGTCTCAGACATCAGACAGCGCTCAGT-----GTGATGAAAGGCTTGCCAGC 1401
QY 361 GLYAspIysValValValGluGlyIleSerIleAlaGlyIleThrGlyAlaIysVal 380
DB 1402 GTGACACAGGTGTGTGTGTCGACGCGCATCTCG-----CGGCTG 1437
QY 381 -ThrProIysGluIrrAlaSerSerGluAsnGlnAlaAla 394
DB 1438 ATGCCCGCGCGCTGTGTCATTCAGAGCGCGCGCGCGC 1480
```

RESULT 19

US-09-199-637A-148
Sequence 148, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
FILE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199, 637A
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066, 517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 148
LENGTH: 1008
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-148

Alignment Scores:

Pred. No.:	4.76e-31	Length:	1008
Score:	370.50 <td>Matches:</td> <td>80</td>	Matches:	80
Percent Similarity:	58.29%	Conservative:	43
Best Local Similarity:	37.91%	Mismatches:	73
Query Match:	18.35%	Indels:	15
	4	Gaps:	5

US-09-889-756a-2 (1-412) x US-09-199-637A-148 (1-1008)

```
QY 205 GLYAspThrThrValIeuAlaThrIleArgGlnThrAsnProMetIrrValAsnValThr 224
DB 1 GGCACGCGCAACCGCATGCGCACCGTGCMAACGCTCGACCCGATCTACGTCACGTCACC 60
QY 225 GlnSerAlaSerGluValMetIysIeuArgArgGlnIleAlaGluGlyIysIeu---Leu 243
DB 61 CAGCGGTCCACCGCGCTGTGCGCATGCGCGCGCAACTGCGCGCGCGCATGTTGAGCGCGC 120
QY 244 ALAlaAspGlyValIleAlaValGlyIleIysPheAspAspGlyThrValTyrProGlu 263
DB 121 GCGCGGACCAACGCTGCGCAAGTCTCCCTGAGAGCTGAGAGAGATGACCAATACCGCTG 180
QY 264 LVSGIYArgLeuIeuPheAlaAspProValValaAsnGluSerThrGlyGlnIleThrIeu 283
DB 181 GAAAGCGCGCTCGAATTCCTCCGAGGTTTCCTGTCGACGAAGCACCGCGCTCGGTCAACATC 240
```



```

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1086
LENGTH: 4404
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1086

Alignment Scores:
Pred. No.: 2,41e-28 Length: 4404
Score: 355.00 Matches: 105
Percent Similarity: 49.73% Conservative: 78
Best Local Similarity: 28.53% Mismatches: 168
Query Match: 17.58% Indels: 18
DB: Gaps: 6

US-09-889-756a-2 (1-412) x US-09-252-991A-1086 (1-4404)

QY 18 ValAlaLeuValLeuSerSerCyGlyLysGlyValAspAlaAlaGlnGlyGlnPro 37
Db ||||| : : : : : ||||| : : : : :
QY 131 GTCGCGCCCTTATCGCTGCTGCGCGGAAATCGCGC-----CCG 169
Db ||||| : : : : : ||||| : : : : :
QY 38 AlaGlyArgGlnAlaProAlaProValValGlyValAlaThrValHisProGlnThrVal 57
Db ||||| : : : : : ||||| : : : : :
QY 170 CCGGGGGCGGAGCGCCCGCCCAAGCGTTCCCGCCGAGGTGGTGGTCCCGCGG 229
Db ||||| : : : : : ||||| : : : : :
QY 58 AlaLeuThrValGluLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAla 77
Db ||||| : : : : : ||||| : : : : :
QY 230 ACACCTACGCGCGGATCACCGGCTCGCTGACCGCTGACGAGTGAACCTCGCGCG 289
Db ||||| : : : : : ||||| : : : : :
QY 78 GlnValGlyGlyLeuGlnGlnLysArgLeuPheGlnGlnGlySerTyValArgAlaGly 97
Db ||||| : : : : : ||||| : : : : :
QY 290 CGCTGCGCGCGTACATCCAGCAGCTTACGCTGCGGAGGCGCGCTGTGAGAGAGGCGC 349
Db ||||| : : : : : ||||| : : : : :
QY 98 GlnProLeuTyArgGlnLeuAspSerSerThrTyGlnAlaLeuLeuGlnSerAlaArgAla 117
Db ||||| : : : : : ||||| : : : : :
QY 350 CAGAACTGTTCTTCATCGACCCCGCGGTTCACAGCGGCGGCGGAGATCCCGCCAGGCA 409
Db ||||| : : : : : ||||| : : : : :
QY 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspValArgTyLys 137
Db ||||| : : : : : ||||| : : : : :
QY 410 CGCTGCGCGAGCGCGAGCGCGGCGTGTGCGCGGACCAACACGAGCGCGAG 469
Db ||||| : : : : : ||||| : : : : :
QY 138 ProLeuValAlaAlaGlnAlaValSerArgGlnGlnTyArgAlaAlaValThrAlaLys 157
Db ||||| : : : : : ||||| : : : : :
QY 470 CTGCTGTATGCGCGAGGTCGTGTCGCGGAGGCGCTCGACAGCGCATCGCTCGCGC 529
Db ||||| : : : : : ||||| : : : : :
QY 158 ArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLeuSerAlaGlyLeuAsn 177
Db ||||| : : : : : ||||| : : : : :
QY 530 AATGCCAGCAAGGCGCGAGTCGATCGAGGCGCGCCCTCGACGCGGCGCAACTGGAT 589
Db ||||| : : : : : ||||| : : : : :
QY 178 LeuAsnArgSerArgLysAlaAlaProLysSerArgPheLeuGlnGlnSerLysValSer 197
Db ||||| : : : : : ||||| : : : : :
QY 590 CTGCGCTTACGCGGCGTACGCGACCGATCGCGGCGGTCTCGGCGCATATCCAGTAC 649
Db ||||| : : : : : ||||| : : : : :
QY 198 GlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrLysGlnThrAsn 217
Db ||||| : : : : : ||||| : : : : :
QY 650 GAGGGCACTACGACCAATGGC---GTCAACCGCGCTGACCAAGCATCGTTTCGTGAT 706
Db ||||| : : : : : ||||| : : : : :
QY 218 ProMetTyVal---AsnValThrGlnSerAlaSerGlnValMetLysLeuArgTyGln 237
Db ||||| : : : : : ||||| : : : : :
QY 707 CCGCTAATCGTGAATCTTCGATGTCAGACGACGACCTACCTGACGAGCGCTCGGCGCAGC 766
Db ||||| : : : : : ||||| : : : : :
QY 237 AlaGlnGlyLysLeuLeuAlaAlaAspGlyValLysAlaAlaValGlyLysPheAspAs 257
Db ||||| : : : : : ||||| : : : : :
QY 767 CGCGGAGGAGGAGCGAG---CAGGCGCCAGGGTC---AAGTGGCGTGTCAACCGACGA 822
Db ||||| : : : : : ||||| : : : : :
QY 257 pGlyThrValTyProGlyLysGlyValGlnLeuPheAlaAspProValLysGlnLys 277
Db ||||| : : : : : ||||| : : : : :
QY 823 GTCC-----TATGAGCAAGACAGTCCGCTTCGATTCCTCGCAACCGCCGACCGCGG 876

```

```

QY 277 rThnGlyGlnLeuThrLeuArgAlaAlaValProAsnAspGlnAsnLysLeuMetProG 297
Db ||||| : : : : : ||||| : : : : :
QY 877 GACCGGACAGGTCGCGGGTTCGGGGGTGTGCAACACCGGAGCGGCACTGACGCGCGG 936
Db ||||| : : : : : ||||| : : : : :
QY 297 yLeuTyValArgValLeuMetAspGlnValAlaAlaAspAsnAlaPheValValProG 317
Db ||||| : : : : : ||||| : : : : :
QY 937 GCTGTTCGCAAGGTCAGCTGAGCCGCAAGCCCTCGCGCGGAGTGTGTGCGCGCA 996
Db ||||| : : : : : ||||| : : : : :
QY 317 nGlnAlaVal---ThrArgGlyAlaLysAspThrValMetLysValAsnAlaGlnGly 336
Db ||||| : : : : : ||||| : : : : :
QY 997 CCATTCATCGGACGACCAAGCGCGCGCTATGTGCTGTGTCGACGAGCAAGCAACA 1056
Db ||||| : : : : : ||||| : : : : :
QY 336 yMetGluProArgGlnValThrValAlaGlnGlnGlnTyThrAsnTrpLysValThrse 356
Db ||||| : : : : : ||||| : : : : :
QY 1057 GACCCAGTACCGGCGCGGTGCACTCGGCGCGGATGTGACAGCGCTCGGCGTCCGCCA 1116
Db ||||| : : : : : ||||| : : : : :
QY 356 rGlyLeuLysAspGlyAspLysValValValGlnGlyLysSerLysAlaGlyLysThrG 376
Db ||||| : : : : : ||||| : : : : :
QY 1117 GGGCTTCAGCCGCGGCGGCGCATCGTCTCAAGGCGCTG-----GTCCGCGC 1164
Db ||||| : : : : : ||||| : : : : :
QY 376 yAlaLysLysValThrProLys 383
Db ||||| : : : : : ||||| : : : : :
QY 1165 GGAATGCAATGATCACCGCGC 1186
Db ||||| : : : : : ||||| : : : : :

RESULT 22
US-09-252-991A-14946
Sequence 14946, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196, 136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14946
LENGTH: 1224
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14946

Alignment Scores:
Pred. No.: 1,22e-27 Length: 1224
Score: 340.50 Matches: 104
Percent Similarity: 51.52% Conservative: 66
Best Local Similarity: 31.52% Mismatches: 145
Query Match: 16.86% Indels: 15
DB: Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-14946 (1-1224)

QY 91 GlySerTyValArgAlaGlyGlnProLeuTyArgGlnLysAspSerSerThrTyGlnAla 110
Db ||||| : : : : : ||||| : : : : :
QY 2 GGGGCACTGTGTAAGAAAGCGACCTGCTTCCATGCAACCGCGCGCTTCGAGGCC 61
Db ||||| : : : : : ||||| : : : : :
QY 111 AsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaAlaThrLeuAlaLysAlaAsp 130
Db ||||| : : : : : ||||| : : : : :
QY 62 GAGGTACAGCGCTCGAAGCCGAGGTGCAACAGCGCGCGCGCCAGCGCGGAGCGTC 121
Db ||||| : : : : : ||||| : : : : :
QY 131 AlaAspLeuAlaArgTyLysProLeuValAlaAlaGlnAlaValSerArgGlnGlnTy 150
Db ||||| : : : : : ||||| : : : : :
QY 122 AACGAAGCCAGGCGCGCAAGCGCTGCGCGGACGAGCGGATCTCCGGAATTCGCC 181
Db ||||| : : : : : ||||| : : : : :
QY 151 AspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAla 170
Db ||||| : : : : : ||||| : : : : :
QY 182 GACGCGCGGACCAACCGCCCGCAGGAAGCGCGGTCGCGCGCGACCGCGCGCA 241

```

```

Qy 171 ILeysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 242 CTGACCGGGCGCGCCCTTAACCTTACCTGAGCTTCAACCGGATCAACGGCGGATGAGCGTGC 301
Qy 191 ILeGlyGlnSerIleValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 302 GTCAGCGCGCGCGGATCACCGCGGCACTGGTCACTCGGGGAA--ACCTGCTC 358
Qy 211 AlaThrIleArgGlnThrAsnProMetCysValAsnValThrGlnSerAlaSerGluVal 230
Db 359 ACCACCTCGTCAACCGACCGAAGGTCTACCTCACTTCGACGGCGGACGAGCGGTGTC 418
Qy 231 MetIys-----LeuArgArgGlnIleAlaGlyIleLeuLeuAlaAlaAspGly 247
Db 419 CTCAGTACGTCGAGCTGGCCCGCAG-----GCCGGTCGCGACACGCGGACGAGAGC 472
Qy 248 ValIleAlaValAlaGlyIleValPheAspAspGlyThrValTyrProGlyIleValArgLeu 267
Db 473 CCGGTCTACCTCGCGCTGAGCAGCGAGACGCGAAC-----CCGACCTGGCGCGGCTG 526
Qy 268 LeuPheAlaAspProValValAlaGlnIleThrGlyGlnIleThrLeuArgAlaAlaVal 287
Db 527 GACTTCCTCGACCAACGAGTCAACCGCGTACCGGACCACTCCGCGCGCGCGCTGTC 586
Qy 288 ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnVal 307
Db 587 GACAACGCGCAGGGCGAGTTCACCCCGGCTCAACGCGCTGAGTGTGTCGCGACG 646
Qy 308 AlaValAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgGlyAlaValAsp 326
Db 647 AAGACTACGCGCGCACCTCATCAGAGCAGAGCGTGGCAGCTGGGCGAGAG 706
Qy 327 ThrValMetIleValAlaGlnIleGlyIleMetGluProArgGluValThrValAlaGln 346
Db 707 TTCTGCTGCTGCTGCTGATGCGCAGACAGACCGTCAACCGGACCGTGAATGGAGCG 766
Qy 347 GlnGlnGlyThrAsnTyrPheValIleThrSerGlyLeuValAspGlyAspValValVal 366
Db 767 AAGTGGAGGGCGCTGCGCATCTGCGCAGCGCGCTGACAGAGGGCGACCGGATCGTGTG 826
Qy 367 GluGlyIleSerIleAlaGlyIleThrGlyAlaValValThrProIys-----Glu 384
Db 827 AATGCGCTGCGAG-----CGGCTCGCGCGCATGAGTGCAGTCCGACAGAGTTCAG 880
Qy 385 TrpAlaSerSerGluAsnGlnAla-----AlaAlaProGlnSerGlyValAlaGlnThr 402
Db 881 ATGGCCAGCGCGCAGACCCCTGGCCACCTCGCGCGCTGCGGCACTGCTGGCGACAGC 940
Qy 403 SerGluAlaValThrAlaSerGluAlaGlu 412
Db 941 GATCCACCGAGGTGGCGCGCTCCAGAGC 970

```

RESULT 23

US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

```

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

```

Alignment Scores:

```

Pred. No.: 1,13e-16 Length: 1830121
Score: 286.00 Matches: 105
Percent Similarity: 41.01% Conservative: 66
Best Local Similarity: 25.18% Mismatches: 148
Query Match: 14.17% Indels: 98
DB: 4 Gaps: 14

```

US-09-889-756a-2 (1-412) x US-09-557-884-1 (1-1830121)

```

Qy 6 PheIysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCys 25
Db 946822 TTTAATATGATTAAGGCGTATGATTAAGCGGAGCCATTGCA----- 946863
Qy 26 GlyIysGlyIleAspAlaAlaGlnIleGlyIleProAlaGlyArgGluAlaProAlaPro 45
Db 946864 -----GGAATGCCA-----GATCTTCAAGCCA 946887
Qy 46 ValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGly 65
Db 946888 GTAACC---GCACCTGAAGTTCAACCGCGTAATGACGCGCATTTAATACAAACAGT 946944
Qy 66 ArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnIys 85
Db 946945 CTGTGGTCCCAATCAAGGCGCATGCTCAGTACACAAATGCGGCGGCTTCACAA 947004
Qy 86 ArgLeuPheGlnGlnIleGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSer 105
Db 947005 GTACTTGTCAAAATGACAAATATGTAAAGGTTAGAGTGTCTTGAGACTTGATAGT 947064
Qy 106 SerThrTyrGluAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
Db 947065 TCTGTGAAGAGCTATCTACAGCTGTCAGGCAACATTAACAGACACTTCGCAACT 947124
Qy 126 LeuAlaValAlaAspAlaAspLeuAlaArgTyrIleProLeuValAlaAlaGluAlaVal 145
Db 947125 -----TACCAACGTTATGTGGCTTATTAATTAACCAATGCTGTA 947163
Qy 146 SerArgGlnGlnIleTyrAspAlaAlaValThrAlaValAspSerAlaGlnAlaGlyValIys 165
Db 947164 TCAGGTCAAGAAATGATTAACGCAAAAGCGCTTATATGATCTCAAGTAGCTAGTATTGAA 947223
Qy 166 AlaAlaGlnAlaAlaIleIysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
Db 947224 TCTCTAAACAGCAATTGA-----CGTCGTAATATTGTCGC 947262
Qy 186 ProIleSerGlyPheIleGlyGlnSerIleValSerGlyIleThrLeuLeuSerAlaGly 205
Db 947263 CCATTGATGCGCAAGCAGGATTTGTGAATAATCATGTTGACAAATATGTGAATGTTGA 947322

```

```

Qy 206 AspThrValLeuAlaThrIleArgGlnThrAsnProMetCysValAsnValThrGln 225
Db 947223 -----ACAGAAATGTGCGGTAGAGATCTAGCTCAATGAAGTGGATTGGCTCTT 947376
Qy 226 SerIAspSerGluValMetLeuLeuArgArgGlnIleAlaGluGlyLysLeuLeuAla 245
Db 947377 TCACAAATATGATTAAGTAAATTAATCAT-----ATCGGTACAGCGGTACAGCG 947424
Qy 246 AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProGluLysGly 265
Db 947425 ACAACAGATGCTCGCTGGGGGAAACATTT-----TCAGCT 947460
Qy 266 ArgLeuPhePheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAla 285
Db 947461 CGAATCACTGGCATTTGAACCTGCCATTATTCATCAACAGCTTTAGTTAGTTACAGCT 947520
Qy 286 AlaVal---ProAspAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeuMet 304
Db 947521 ACTTTGATCTCGAAGATGGGCAATAAATTGCTTCAGGTATGTTCTCTCGCTTACGAT 947580
Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr----- 321
Db 947581 GCACCTTCCAACCTGAACAAATCAAGTTGTCGTCCACAGTACATTAGCTACAAATATG 947640
Qy 321 ----- 321
Db 947641 TATGGCAAAATTCCTATTACTTGAACCATTTCTGAAGAAGAAAAAGAAATATGCA 947700
Qy 322 -----ArgGlyAlaLysAspThrValMetIleValAsn 332
Db 947701 GGTAATGAAATAATGATCGTCTGATCGTCGAAACAGATCACCGATTAATTAAGAT 947760
Qy 333 AlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrp 352
Db 947761 CGTAAAGT-----GTTATATGCTCAATTACAGGGAAATGAA--- 947796
Qy 353 IleValThrSerGlyLeuLysAspGlyAspLysValValIleGluGlyIleSerIleAla 372
Db 947797 -----GTTAAAGTGGAGATTAATTAATTACAGCGGTACAGAA----- 947835
Qy 373 GlyIleThrGlyAlaLysLysValThrProLysGluTrpAlaSerSerGlu 389
Db 947836 GGTAATGTAATGGAAGCTGTGTG-----CAATGGATTAAAAAAGAC 947877

RESULT 24
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A

```

```

FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 1,13e-16 Length: 1830121
Score: 286.00 Matches: 105
Percent Similarity: 41.01% Conservative: 66
Best Local Similarity: 25.18% Mismatches: 148
Query Match: 14.17% Indels: 98
Gaps: 14

US-09-889-756a-2 (1-412) x US-09-643-990A-1 (1-1830121)

Qy 6 PheLysAlaMetArgAlaAlaIleValAlaValAlaValLeuValLeuSerSerCys 25
Db 946822 TTTAATATGATTAAGGCGTAATGATTAAGCGGACCATTTGCA----- 946863
Qy 26 GlyLysGlyLysPheAlaIleGlnGlyGlnProAlaGlyArgGluAlaProAlaPro 45
Db 946864 -----GGAATGCCA-----GAAATCTCAAGCCCA 946887
Qy 46 ValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGly 65
Db 946888 GTAAAC---GCACCTGAAGTTCAACCGCGTAATGACGCGCATTTAACAACAAGGT 946944
Qy 66 ArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLys 85
Db 946945 CTGTGCGTCGAATCAAGCGCAATGCTCAGTACACAAATGCGCGCTTCACAA 947004
Qy 86 ArgLeuPheGlnGluGlySerTyValArgAlaGlyGlnProLeuTyArgIleAspSer 105
Db 947005 GTACTGTTCAAATGACAAATATGTAAAGGTAGGTGCTTGGAGCTTGATAGT 947064
Qy 106 SerThrTyGlnAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
Db 947065 TCGTTTAAACGACTATCTACAAGCTGCTCAGGCAAAATTAACAGACTTCGTAAC 947124
Qy 126 LeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeuValAlaAlaGluAlaVal 145
Db 947125 -----TACCAACGTTAATGTGGGTTAATTAATACCAATGCTGTA 947163
Qy 146 SerArgGlnGluTyraAspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLys 165
Db 947164 TCACGTCAGAAATATGATTAAGCAAAAGCGCTTAATGATCTCAAGTACAGTATGAA 947223
Qy 166 AlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
Db 947224 TCTCTAAAGCACCAATTGAA-----CGTCAATAATTTGTTGCG 947262
Qy 186 ProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThrLeuLeuAsnAlaGly 205
Db 947263 CCATTGTATGCAAGACAGGTATTTGTGAATAATCAATGTTGACAAATATGTGAATGTTGGA 947322

```

```

OY 206 AsphtrThrValIleAlaThrIleArgGlnThrAsnProMetTyrValAlaValThrGln 225
Db 947323 -----ACAGAAATGTCGCGTTCAGACAGATCTACTCAATGAAGATGGAATTTGCTCTT 947376
OY 226 SerAlaSerGluValMetLeuLeuArgArgGlnIleAlaGluGlyLeuLeuAlaIle 245
Db 947377 TCACAAATGATTATGATTAATTCAT-----ATCGGTACACGGCGTTACACGG 947424
OY 246 AspGlyValIleAlaValGlyIleLeuSphaSerPheGlyThrValTyrProGluLeuSly 265
Db 947425 ACAACAGATGCTCGCTTGGCGCAACATTT-----TCACCT 947460
OY 266 ArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeuArgAla 285
Db 947461 CGAATCACTGCGATGAACCTGCGCATTAATCATCAACAGGTTTATGATGTTACAGGCT 947520
OY 286 AlaVal-----ProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMet 304
Db 947521 ACTTTTGATCCTGAAGATGGCATTAATGCTTCAGGATGATGTTCTTCGCTTACGCAAT 947580
OY 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThr----- 321
Db 947581 GCACCTCCACAGTGAACAAACATCAAGTTGCTTCACAGTACGATTAAGCTACAAATATG 947640
OY 321 ----- 321
Db 947641 TATGCGCAATGCTCTATTACTTGACCATTAATCTGAAGAAAGAAAGAAATGTCATCA 947700
OY 322 -----ArgGlyAlaValAspThrThrValMetIleValAsn 332
Db 947701 GGTAAATGAAAAATGATGATGCTCTATCGTGCAGAAACAGATCAACGATTTTAAATGAT 947760
OY 333 AlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsnTyr 352
Db 947761 CGTCAAGT-----GTTATGCTCAATTAACAGGAAATGAA----- 947796
OY 353 IleValThrSerGlyLeuLysAspGlyAspLysValValValGluGlyIleSerIleAla 372
Db 947797 -----GTTAAAGTGGAGATTAATAATTATTACAGCGGTCACAGAA----- 947835
OY 373 GlyIleThrGlyAlaValLysValThrProLysGluTTPalaseSerGlu 389
Db 947836 GGTATGTTGTAATGGAAGCTTGTG-----GAATGATTAATAAAGAC 947877

RESULT 25
US-09-252-991A-7914
; Sequence 7914, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7914
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7914

Alignment Scores:
Pred. No.: 9,94e-21 Length: 1185
Score: 277.00 Matches: 104
Percent Similarity: 45.57% Conservativeness: 71
Best Local Similarity: 27.08% Mismatches: 141
Query Match: 13.72% Indels: 68
DB: 4 Gaps: 14

```

```

US-09-889-756A-2 (1-412) x US-09-252-991A-7914 (1-1185)
OY 14 LeuAlaIleAlaValAlaLeuValLeuSerSerGlyGly-----LysGlyGlyAspAlaIle 32
Db 82 CTCGCCGGGATGATGCGCGTGTGGCATTTCTCCCGGCTACAGAGGCTTACTCATTCCT 141
OY 33 GlnGlyGlyGlnProAlaGlyArgGluAlaProAlaProValValGlyValThrVal 52
Db 142 CAG-----CAGATCGCCCTTTTCAGGCAACCGAATCGGCATCAGCGTG----- 186
OY 53 HisProGlnThrValAlaLeuThrValGluLeuPro----- 64
Db 187 -----ACCGCACCTCGCGCGAAAGACGTCCTCGGACAGACCCCTCCGACCATC 237
OY 65 GlyArgLeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGln 84
Db 238 GGCAGCCTCAAGGATTCACAGGCGGTGACCTTCAACCCCGAAGTCTCCGACGCGTAC 297
OY 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAsp 104
Db 298 GACGTACTGTTCTTCCGGGACACAGTACAGTGAAGTGAACACCGTATCAGTGGAA 357
OY 105 SerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124
Db 358 AGCAGCGTCAGAGAGACACCTCGCG-----ACTGCGAGGCG 396
OY 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGluAla 144
Db 397 GATCTCGCGCTGCGCGACGCGGATACAGCGCGCGCGCACTGATCGGACGAGAGCC 456
OY 145 ValSerArgGlnGlyTyrAsp-----AlaAlaValThrAlaLysArgSerAlaGluAla 162
Db 457 ATCTGAAAGCGAATTCGATCGTCTCGCGCGAGGCGCAAGACAGACGACCGCATC 516
OY 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 517 -----GCCGAGCTGAGAGCGCGCG-----CTGCGAAGAGAGACGCG 549
OY 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerTyrValSerGluGlyThrLeu 202
Db 550 GTGCTCGCGCTTCGCGCGGACCATCGCATCGCGACGATGACGTCGCGACTAC 609
OY 203 AsnAlaGlyAspThrThrValIleAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db 610 TCGCGCGCG-----ACGCCGATCGCACCTTCGACAGCACTTCCACCTCGCTCGAT 663
OY 223 ValThrGlnSerAlaSerGluValMetLysLeuArgArg--GlnIleAlaGluGlyLys 241
Db 664 TTCACCTGCGCGACGACGACTTCCCTCTCAGCGCGCGACGCTGTGAAGTCCG 723
OY 242 LeuLeuAlaIleAspGlyValIleAlaValGlyIleLysPheAspArgIleThrValTyr 261
Db 724 GTCCCGCGCTACCCCGCGCGGTG-----TTCCAGC----- 753
OY 262 ProGluLysGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIle 281
Db 754 -----GCCAGATCGCGCGCATCAACCCAGGTGACACAGACCGCGAACCTG 804
OY 282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal--- 300
Db 805 CAGGTCCGCGCTGCTCGGAGAACCCGACGCAAGCTGCTCGCGGCGCATGTCGCCAAC 864
OY 301 -----ArgValLeuMetAspGlnValAlaVal--- 309
Db 865 CTCGAGGTATGTTGCTGCGCGAGAACCAACAGCTGCTGCGCGGACGCGCATACC 924
OY 310 -----AspAlaAlaPheValValProGlnGlnAlaValThrArgGlyAla 324
Db 925 TTCACCTTACGCGCATCTGATCTGCTGCGGAGAAAGACGACGACGCGCAG 984
OY 325 LysAspThrValMetIleValAsnAlaGlnGlyMetGluProArgGluValThrVal 344

```



```

Db 406 GAGCTGATCCGCTCAAGCCGAGTTGCGCAATGCCGAGATCTTCATGCCCGCGC--- 462
Qy 131 AlaaspleuAlaArgTyrIleuProleuValAlaIaIaValaIaValaSerArgIngluTyr 150
Db 463 -----CGCAAGCTGGTAGAGCCCAACGTGCGCTCCAGAGAACAGCTG 504
Qy 151 AspaIaIaValaIaIaValaIaIaValaIaIaValaIaIaValaIaIaIaIaIaIaIaIa 170
Db 505 GACAAAGCGCTGCGCGCGCGAGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 564
Qy 171 IleuSerArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 190
Db 565 ATC-----GACCAAGAGCGGATCCGCGCGCGCTTCCTCCGCGAG 603
Qy 191 IleuGlnSerIleuValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 210
Db 604 CTCGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 663
Qy 211 AlaThrIleArgInlThraIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 230
Db 664 CTGGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
Qy 231 MetIleuSerArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 250
Db 724 CTGAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Qy 251 ValGlyIleuValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 270
Db 748 GTCCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 801
Qy 271 AspProValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 290
Db 802 GACCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858
Qy 291 GlnAsnIleuMetProGlyIleuTyrValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 310
Db 859 GAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 918
Qy 311 AsnaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 329
Db 919 CCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 978
Qy 330 IleValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 349
Db 979 GTCCCGCAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1038
Qy 350 ThrAsnTrp-----IleValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 365
Db 1039 GAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1098
Qy 366 ValGluGly 368
Db 1099 ACTTCCGGA 1107

RESULT 29
US-09-252-991A-4005
; Sequence 4005, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4005
; LENGTH: 1398
; TYPE: DNA

```

```

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4005

Alignment Scores:
Pred. No.: 1,596-19 Length: 1398
Score: 267.00 Matches: 93
Percent Similarity: 43.25% Conservative: 64
Best Local Similarity: 25.62% Mismatches: 170
Query Match: 13.22% Indels: 36
DB: 4 Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-4005 (1-1398)

Qy 11 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 30
Db 32 AGTCCCGCGCTCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
Qy 31 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 50
Db 92 AAGAAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 145
Qy 51 ThrValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 70
Db 146 TCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205
Qy 71 ArgThrIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 90
Db 206 GCGCAGGTCCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 265
Qy 91 GlySerTyrValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 110
Db 266 GCGCAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 325
Qy 111 AsnIleuSerIleuValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 130
Db 326 GAGCTGATCCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 382
Qy 131 AlaaspleuAlaArgTyrIleuProleuValAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 150
Db 383 -----CGCAAGCTGGTAGAGCCCAACGTGCGCTCCAGAGAACAGCTG 424
Qy 151 AspaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 170
Db 425 GACAAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 484
Qy 171 IleuSerArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 190
Db 485 ATC-----GACCAAGAGCGGATCCGCGCGCGCTTCCTCCGCGAG 523
Qy 191 IleuGlnSerIleuValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 210
Db 524 CTCGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 583
Qy 211 AlaThrIleArgInlThraIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 230
Db 584 CTGGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 643
Qy 231 MetIleuSerArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 250
Db 644 CTGAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 667
Qy 251 ValGlyIleuValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 270
Db 668 GTCTCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 721
Qy 271 AspProValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 290
Db 722 GACCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778
Qy 291 GlnAsnIleuMetProGlyIleuTyrValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 310
Db 779 GAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838

```



```

QY 311 AsnAlaPheValProGlnGlnAlaValThrArgGlyAla---LysAspThrValMet 329
Db 839 CCGTCCGTCGACCGTCGCGGAACCGCGTCACTATACCGGTACCGGACACCGGTTC 898
QY 330 IleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGly 349
Db 899 GTCCCCACACGACGACGCGGACCGCGCTCAGCCGCAAGCGGTCTCGGTCCGATCCGC 958
QY 350 ThrAsnTrp-----IleValThrSerGlyLeuLysAspGlyAspLysValVal 365
Db 959 GAGCGCTGGAGACGCGTCCGTTGAAATCTCCAGGCGCTCGCCGAGGCGACCGGTATG 1018
QY 366 ValGlnGly 368
Db 1019 ACTCCGGA 1027

RESULT 30
US-09-252-991A-3906/c
; Sequence 3906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3906
; LENGTH: 4704
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3906

Alignment Scores:
Pred. No.: 1,08e-18 Length: 4704
Score: 267.00 Matches: 93
Percent Similarity: 43.25% Conservative: 64
Best Local Similarity: 25.62% Mismatches: 170
Query Match: 13.22% Indels: 36
DB: Gaps: 8

US-09-889-756a-2 (1-412) x US-09-252-991A-3906 (1-4704)
QY 11 AlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyGlyGlyAsp 30
Db 4500 AGTCCCGGCTCATCTGCGCGCGGCGGTATCGCATCCCGCTACGCCACCGCGCTCGGCG 4441
QY 31 AlaAlaGlnGlyGlyGlnProAlaGlyArgGluAlaProAlaProValAlaGlyValVal 50
Db 4440 AAGAAAGACGCGCGGCTTTCGCGGC-----TACCGCGCGTAAGGTGCGCTCGCC 4387
QY 51 ThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGlnSerLeu 70
Db 4386 TCGGTGAGCGCGCGGTGTCGCGCGCTCTTCATGCGGTGCGGTGAGCGCGGT 4327
QY 71 ArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlu 90
Db 4326 CGCCAGGTCCAGGTGCGCGCGGACGCGGACGATCACCAGATCGCTTCGAATCG 4267
QY 91 GlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAla 110
Db 4266 GCGCAGCAGGTGCGGACGAGGCGATGCTGTGCGCACTCAACGACGCGGTGGAACAGCC 4207
QY 111 AsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAsp 130
Db 4206 GAGCTGATCCCTCTCAAGGCGCGATGTGCGCAATGCGGATCTCCATGCGCGCG-- 4150
QY 131 AlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnValAlaSerArgGlnGlnTyr 150

```

```

Db 4149 -----CCAAAGCTGTAGAGCGCAACGTCGCTCGCAGAACACTG 4108
QY 151 AspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaAlaAla 170
Db 4107 GACAAACCGCTGCGCGCGCGGACGATGCGCTCGCGCGCGCGCACAGACCCAGCGCTG 4048
QY 171 IleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 4047 ATC-----GACCAAGGCGATTCGCGCGCGCTTCTCGGCGAG 4009
QY 191 IleGlyGlnSerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 4008 CTCGCGATCCGCGCGCGCGGACCTTCGCGAGTCTCGCGCGCGCGCGCGCGCGAGC 3949
QY 211 AlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluVal 230
Db 3948 CTGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 3889
QY 231 MetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
Db 3888 CTGAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 3865
QY 251 ValGlyIleLysPheAspAspGlyThrValTyrProGlyLysGlyArgLeuLeuPheAla 270
Db 3864 GTCTGTGTCGACGCGCTATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 3811
QY 271 AspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
Db 3810 GACCGCGTGAATCGCGCAAGTGGCGCG--ACGTCGAGGTCCAGGCGCTTGCGGACACCCC 3754
QY 291 GlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAsp 310
Db 3753 GAAAGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3654
QY 311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAla---LysAspThrValMet 329
Db 3653 CCGTCCGTCGACCGTCGCGCGGACGCGGATCGCATTCACCGCTACCGCGCGACACCGGTTC 3634
QY 330 IleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGly 349
Db 3633 GTCCCCACACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3574
QY 350 ThrAsnTrp-----IleValThrSerGlyLeuLysAspGlyAspLysValVal 365
Db 3573 GAGCGCTGGAGCGTCCCGTGAATCTCCAGGCGCTCGCCGAGGCGACCGGCTACTG 3514
QY 366 ValGlnGly 368
Db 3513 ACTCCGGA 3505

RESULT 31
US-09-252-991A-12042/c
; Sequence 12042, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12042
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12042

```


Db 778 ---CTGCTCGGCAAGCCGGAAGTCAACGGC----- 804

Qy 261 TyrProGluLysGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGln 280

Db 805 -----AGCGGCAAGAGTCCGCGAGATCAACCCGAGGTGAGACGACGCGGCTACG 855

Qy 281 IleThrLeuAlaGlnA-----AlaValProAsnAspGlnAsnIleLeuMetProGly 297

Db 856 CTGAAGGTCAGAGGTGGCGCTCGACTCGGTGGCGGCGGAAATGAGCTC-----GGC 906

Qy 298 LeuTyrAlaGlyValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGln 317

Db 907 AGCGTGTCAACGCCAGCGTCCGCCCGCGCGC-----GAGCAAGCGGTGTCTGCTCG 963

Qy 318 GlnAlaValThrArg---GlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGly 336

Db 964 TCGGGCGGTCTCAAGGTGGCGGAGCGCGCGGTCTGTCTCGACACGACGACGAG 1023

Qy 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrPheValThrSer 356

Db 1024 GCGCGTCTGCAACCGGTGCGGTGGCGCTACGCCAGACGAGAAGTGTATCGACGCT 1083

Qy 357 GlyLeuLysAspGlyAspLysValValValGluGly 368

Db 1084 GGCGTGAAGCGCGCCAGACGCTGTCTACGGTGGCG 1119

RESULT 34

US-09-252-991A-7845

Sequence 7845, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252, 991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 7845

LENGTH: 1011

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7845

Alignment Scores:

Pred. No.: 2,85e-18 Length: 1011

Score: 253.50 Matches: 90

Percent Similarity: 46.56% Conservative: 59

Best Local Similarity: 28.12% Mismatches: 120

Query Match: 12.56% Indels: 51

DB: 4 Gaps: 12

US-09-889-756a-2 (1-412) x US-09-252-991A-7845 (1-1011)

Qy 14 LeuAlaAlaAlaValAlaLeuValLeuSerSerCysGly---LysGlyGlyAspAlaAla 32

Db 95 CTGCGCCCGGTGATGCGCGGTGTGCGATTCTCGCCGCTACAAAGGTCTACTCCATCCGT 154

Qy 33 GlnGlyGlyGlnProAlaGlyArgGluAlaProAlaProValValGlyValValThrVal 52

Db 155 CAG-----CAGATCGCCCTTTTTCAGCGCACCGAAACCGCGATCAGCGTG----- 199

Qy 53 HisProGlnThrValAlaLeuThrValGluLeuPro----- 64

Db 200 -----ACCGCAGCCTGTGCGCGAAAGAGTCTCTGGGAGAGCGCCCTGCCAGCCATC 250

Qy 65 GlyArgLeuGluSerLeuArgThrAlaAspValAlaArgAlaGlnValGlyGlyIleLeuGln 84

Db 251 GGAGGCTCAAGGATTCACAGGGCGTGACCTCACCGCCGAAAGTCTCCGACGCTACG 310

Qy 85 LysArgLeuPheGlnGlnGlySerTyrValArgAlaGlnProLeuTyrGlnIleAsp 104

Db 311 GACGACTGTTCTTCCCGGAGCAGAGTGAAGCTGACCAACCGCTGATCAGTTGGAA 370

Qy 105 SerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAla 124

Db 371 AGCAGCTCGAGAGGACCCCTTCGCG-----ACTGCCAGAGCC 409

Qy 125 ThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAla 144

Db 410 GATTCGCGCTCGCGCAGGCGCAGATACAGCGCGCGCGCAACTGATCGCAGAGGCC 469

Qy 145 ValSerArgGlnGlnTyrAsp-----AlaAlaValThrAlaLysSerAlaGlnAla 162

Db 470 ATCTGAAAAGCGAATTCGATGCTGTCTGCGCGCGAGAGGCGCAAGACGAGCCACCGTC 529

Qy 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182

Db 530 -----GCCAGCTGAAGCGCG-----CTGGCGAAGAGCGC 562

Qy 183 IleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlyThrLeuLeu 202

Db 563 GTGCTCGCGCTTTCGCGGAGCACCATCGCATCGCCAGGTGACGTGCGGACTACGTC 622

Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222

Db 623 TCGCGCGCG-----ACGCGCATCGCACTTGACAGACCTTTCACCTGCTCTCGAT 676

Qy 223 ValThrGlnSerAlaSerGluValMetLysLeuArgArg---GlnIleAlaGlnLys 241

Db 677 TTCACCTGCGCCAGCAGACGATTCCTCCCTCGATCAGCGCGCGAGCTGTGGAAGTCCGG 736

Qy 242 LeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgGlyThrValTyr 261

Db 737 GTGCGCGCTTACCCCGCGCAGGTG-----TTCGAC----- 766

Qy 262 ProGluLysGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIle 281

Db 767 -----GCCAGATCGCGCGCCATCAACCCCAAGTTCACACAGACGACCGACCTG 817

Qy 282 ThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArg 301

Db 818 CAGGTCCGCGCTGCTCGGAGAACCCGAGCGCAAGCTGCTGCGGCGATGTTGCCAAC 877

Qy 302 ValLeuMetAspGlnValAlaValaAspAsnAlaPheValValProGlnGlnAlaValThr 321

Db 878 CTCGAGGTGATGTTGCTGCGCGAGAAACAACGCTGTGTGCGCGAGACGCGATCAC 937

RESULT 35

US-09-252-991A-4631

Sequence 4631, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252, 991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 4631

LENGTH: 1170

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-4631

Alignment Scores:

Pred. No.: 2.68e-17 Length: 1170

Db 296 CTCGATCCCGACGAGTCGCGCTCGCACTGAGGCGCGCGCCGACGTCAGTCCGCC 355
Qy 123 GlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAla 142
Db 356 GAGGCCAACTTCGACAGCGTCCGCGGAGTACCGGCTTACCGCACTTCTCTGACCGC 415
Qy 143 GlnAlaValSerArgGlnTyrAspAlaAlaValThrAlaValArgSerAlaGlnAla 162
Db 416 AACCTGGTCACCACTTCCAGTTCGAGAACATCCAGAACACTACCGCGCGGAGGCGC 475
Qy 163 GlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArg 182
Db 476 CGGTGAAGACGATCCCGCGCAATTCACCTCCCGCAACACGCGCGGTACGCGTG 535
Qy 183 IleThrAlaProLysSerGlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeu 202
Db 536 CTGGCTTCGCGCCGACGATGCGTATCCCGCGCGCGCGCTCGAGGTGGCCAGGTGGT 595
Qy 203 AsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsn 222
Db 596 GCGGCGCGACGACGCTTCAGCCTGGCGCGCGACGCGCAACGCGAGTCTGATCGGC 655
Qy 223 ValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeu 242
Db 656 CTGCGGAAACACACTTCGACGTTCTCGCATCGCGCACCGCGGTGCGTCACTCTGG 715
Qy 243 LeuAlaAlaAspGlyValIleAlaValIleLysPheAspAspGlyThrValTyrPro 262
Db 716 TCGCAAGCGCACAGCCTTCGCGCGGATTCGCGAGCTCTGCGCGCGCGCGATCG 775
Qy 263 GlyLys-----GlyArgLeuLeuPheAlaAspProValAlaAsnGlnSerThr 278
Db 776 CAATCGGTACTTCGCGCGCGGAGTGGCTTCGACACACCGCGGACCTCGCGCGAATCG 835
Qy 279 GlyLys-----IleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 836 GCGCGAGCGCGCGGTCTACGTCGCGCGCGCGCGCGCGCGGTGCGCG----- 880
Qy 294 LeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 881 -----TTA 883
Qy 314 ValValProGlnGlnAlaValThrArgGlyValAlaLysAspThrValMetIleValAsnAla 333
Db 884 TCGGTTCTCTGTCGCGGTACCGCACAGCGCGCGCGCGGTCTGCTGGGTG----- 937
Qy 334 GlnGlyMetGluProArgGlnValThrValAlaGlnGln----- 347
Db 938 -----GTGACGCGCGCGCGCTCGCTCGCGCGCGCGGTGCGCACCGGTGCC 988
Qy 348 ---GlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValVal 366
Db 989 TATGCCAGAGACCGGTCGCGGTCTCGAAGGCTGAAGGCTGAGCTGGGTGGTGGCC 1048
Qy 367 GlnGlyLysSerIleAlaGlyIleThrGlyAlaLysValThrProLysGlnTyrAla 386
Db 1049 ACGGGGTCAAGT-----CTTCGGAAGGAGGACAGAGTGGTCCGATCGACCGGGCC 1102
Qy 387 SerSerGlnAsnGlnAlaAlaAlaProGln 396
Db 1103 AACCGACAGGTGAACCTGCGCGCGCAAGAG 1132

RESULT 37
US-09-252-991A-14157/c
; Sequence 14157, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14157
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14157

Alignment Scores:
Pred. No.: 1,16e-17 Length: 588
Score: 244.50 Matches: 56
Percent Similarity: 56.83% Conservative: 23
Best Local Similarity: 40.29% Mismatches: 55
Query Match: 12.11% Indels: 5
Gaps: 3

US-09-889-756a-2 (1-412) x US-09-252-991A-14157 (1-588)

Qy 261 TyrProGlnLysGlyValLeuLeuPheAlaAspProValValAsnGlnSerThrGlyLys 280
Db 587 TACGAGCGCCAGGCGCGGTGCGAGTTCGCGAGTGGCGGTGATCGCGGTACCGGCCAG 528
Qy 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
Db 527 ATGCGCTTCGCGCGCAAGTTCGCGCAACCCGCGAGGGTCTCTGCTCGCGCGCATGTACGTG 468
Qy 301 ArgValLeuMetAspGlnValAlaValAlaAspAsnAlaPheValValProGlnGlnAlaVal 320
Db 467 CGGTACGTACGCGCCAGCGGATCGACCAACCGAGTGTGGTGGCGCGCGCGCGGTG 408
Qy 321 ThrArgGlyAlaLysAspThr-----ValMetIleValAsnAlaGlnGlyMetGlu 338
Db 407 CACCGCTTCGCGCGCACG 348
Qy 339 ProArgGlnValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeu 358
Db 347 TCGCGACGCTCCGTCACCG 288
Qy 359 LysAspGlyAspLysValValValAlaGlnGlyLysSerIleAlaGlyIleThrGlyAlaLys 378
Db 287 GACCGCGGTGACCGGGTCAATGTCGGCGC-----CTGCTCGCGGTGCGCGCGGGGTG 234
Qy 379 LysValThrProLys---GluTrpAlaSerSerGlnAsnGlnAlaAlaAlaProGln 396
Db 233 AAGATCGTCCGGAAGCGCGATGTGTGCCAGCGCAAGCCAGTCACCTGCGCGCA 177

RESULT 38
US-09-252-991A-11448/c
; Sequence 11448, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11448
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11448

Alignment Scores:
Pred. No.: 3.19e-17 Length: 465


```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9189
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-9189

Alignment Scores:
Pred. No.: 1,41e-14 Length: 1353
Score: 221.50 Matches: 101
Percent Similarity: 40.58% Conservative: 67
Best Local Similarity: 24.40% Mismatches: 184
Query Match: 10.97% Indels: 62
Gaps: 12

US-09-889-756A-2 (1-412) x US-09-252-991A-9189 (1-1353)

QY 35 G1G1nProAlaG1AArgG1uAlaProAlaProAlaVal-G1yValValThrValHisPr 54
Db 1198 GGCCTGCGTGGAGCCATCCGTCCTCCGACACAGCTTCGATCCGTCAGCCGACCG 1139
QY 54 oG1nThrValAlaLeuThrValG1uLeuProG1AArgLeuSerLeuArgThrAlaAs 74
Db 1138 CGGAGATCGAGAGCAGCGCTCGGCGCTCGGACCTCGAACCAGCGGCGTACGTGA 1079
QY 74 pValArgAlaG1nValG1yLeuLeuG1nValArgLeuPheG1nG1ySerTyVa 94
Db 1078 CGTGGCGCCGAGCCCTCGGCGATCCGAGCTGCGAGCCGCGGCGGAGATGT 1019
QY 94 1aGAlaG1yG1nProLeuTyrg1nLeuAspSerSerThrTyG1uAlaLeuG1uSe 114
Db 1018 GAGCGAAGGCGATGTGTGTGTCAGATCGACCTCCACCCAGCAGCCCAAGTCATGC 959
QY 114 rAlaArg-----AlaG1nLeuAlaThrAlaG1nAlaThrLeuAl 127
Db 958 CGCGCGCTATTCATCGATCGATGCTCAAGCGCCAGCTGCGGAGCAAGTCCATAC 899
QY 127 alysAlaAspAlaAspLeuAlaArgTyryspProLeuValAlaAlaG1uAlaValSerAr 147
Db 898 CTTGCGCCCGCAGCAGTACCAAGCGCCAGCGGCTGGCGCGCGCGGAGCGGTAC 839
QY 147 gG1nG1uTyrAspAlaAla-----ValThrAlaAlaArg-----SerAl 160
Db 838 CGAGGAGCTGACAGAGCCGCCAGGCGCAGATGCTGCCACCCAGGCAAGTCCAGATGA 779
QY 160 ag1uAlaG1yVal1yAlaAlaG1nAlaAla1e1ySerAlaG1y1LeuAsnLeuAsnAr 180
Db 778 CAGGCGCAGATCCGCGAGCCAGCGCTCGTGGCGAGCGAAGCCAGAGCTCGGCTA 719
QY 180 gSerArg1LeuThrAlaPro1LeuSerG1yPhe1LeuG1nSer1yValSerG1uG1yTh 200
Db 718 TACCGCGATTAAGCGCGGATGTCGCGCAGCGTGGCGGATGCGCGGAGGCGCA 659
QY 200 rLeuLeuAsnAla---G1yAspThrThrValLeuAlaThr1eArg1nThrAsnProMe 219
Db 658 GACCTCAATGCCAGCAGCAGACAGACCCGTTGATCTCGGATCGCCAAATTGTGCGGAT 599
QY 219 tTyrValAsnValThrG1nSerAlaSerG1uAlaMet1yLeuArgArg1nLe---Al 238
Db 598 GACGCTGTGGCCAGCGATTCGGAAGCGCAGTCGCGGATCAAGCCGCGATGCGCGC 539
QY 238 ag1uG1y1yLeuLeuAlaAlaAspG1y-----Val11 249
Db 538 CTACTTCACACCTTCAGCGGAGAGCGCGGCTGAGACCGGCAAGTCCGCGAGATCCT 479

```

```

QY 249 ea1aValG1y1e1yS-----PheAspAspG1yThrValTyRProG1uTySg1 265
Db 478 CCGGTGCGCCCGCAAGCCGCTGAGCAGCAGCAACAGGCGGCGGAGCCCGCACACGG 419
QY 265 yArgLeuLeuPheAlaAspProValAlaLeuG1uSerThrG1yG1n----- 280
Db 418 C-----AGCGGCGGCGAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 392
QY 281 -----1LeThrLeuArgAlaAlaValProAsnAspG1yAsn1LeuMetProG1 297
Db 391 GGTGTGCTGTATACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 332
QY 297 yLeuTyRValArgValAlaLeuMetAspG1nValAlaValAspAsnAlaPheValValProG1 317
Db 331 AATGACCGCCAGAGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 272
QY 317 nG1nAlaValThrArgG1yAla1yS-----AspThrValMet1LeuValAsnAlaG1nG1 335
Db 271 CGCGGCGATTCACGAGCAGCAGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 212
QY 335 yG1yMetG1uProArgG1uValThrValAlaG1nG1nG1nG1y1yThrAsnTrp1LeuAlTh 355
Db 211 CAAGCCGCGAGCGCGCGATTCAGACCGGCAATCAGCAGCGGCTGCGGCTACAGGTGCT 152
QY 355 rSerG1yLeuTyAspG1yAsp1ySValValAlaValAlaG1uG1y1yLeuSer1LeaG1y1eTh 375
Db 151 TGCCTGCTGTGAGAGAGGAGCGCGCTGTGARG----- 118
QY 375 rG1yAla1y1ySValThrPro1ySg1uTrpAlaSerSerG1uAsnG1nAlaAlaPr 395
Db 117 -GCCGCTCCGAGCGGAGCAGCAGTGTGAGAAACGCCAGCAAC-----CCGTCCC 65
QY 395 oG1nSerG1yAlaG1nThrAlaSerG1uAla1ySThrAla 408
Db 64 CTGATCAACTGCGGACATCCGCAAGCGCTACCGCGGCA 25

```

Search completed: September 8, 2003, 07:58:18
Job time : 2330 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 04:57:54 ; Search time 368 seconds
(without alignments)
3022.197 Million cell updates/sec

Title: US-09-889-756a-2
Perfect score: 2019
Sequence: 1 MAFYAFKAMRAALAAVAL.....AAPQGVQTASEAATSEAE 412

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Xgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 252756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xjh
-Q=/cgr2_1/USFTO.spool/US09889756/runat.05092003.124323.23271/app.query.fasta.1.583
-DB=N.geneseq.19Jun03 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNIT5-bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09889756 @CGN 1.1 312 @runat.05092003.124323.23271 -NCP=6 -ICPU=3
-NO_MMAP -JARBOTERY -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.GeneSeq.19Jun03.*
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2019	100.0	1239	21	AAA37645	N. meningitidis BA
2	1989	98.5	1239	24	ABK54080	DNA encoding Neiss
3	1989	98.5	20352	21	AA81488	N. meningitidis pa
4	1988	98.5	349980	21	AA21611	Neisseria meningit
5	1989	98.5	837096	21	AA81489	N. meningitidis pa
6	1943	96.2	1236	25	AB239062	N. gonorrhoeae nuc
7	762.5	37.8	4879	20	AA211064	E. coli acra and a
8	701.5	34.7	7888	24	AB578861	E. coli CF073 gen
9	631.5	31.3	99629	22	AA28550	Genomic fragment #
10	470	23.3	2967	23	AA589861	DNA encoding novel
11	459.5	22.8	2934	23	AA579793	DNA encoding novel
12	435.5	21.6	1395	21	AA65929	E. coli proliferat
13	435.5	21.6	1395	22	AA84642	E. coli growth and
14	423	21.0	1116	24	AB090066	M. capsulatus gene
15	370.5	18.4	1068	20	AA598220	Nucleotide sequenc
16	339	16.8	2112	23	AA579919	DNA encoding novel
17	339	16.8	2112	23	AA581885	DNA encoding novel
18	324	16.0	2310	23	AA582449	DNA encoding novel
19	313	15.5	1176	25	AB240695	N. gonorrhoeae nuc
20	310	15.4	1179	20	AA212306	Neisseria gonorrh
21	309.5	15.3	349980	21	AA21607	Neisseria meningit
22	309.5	15.3	143768	21	AA81490	N. meningitidis B
23	308	15.3	1176	24	AA597298	Neisseria meningit
24	308	15.3	1177	24	AA597248	Neisseria meningit
25	308	15.3	1179	20	AA212305	Neisseria meningit
26	308	15.3	1179	21	AA253631	Neisseria gonorrh
27	308	15.3	1179	21	AA253633	Neisseria meningit
28	304.5	15.1	3732	23	AA574537	DNA encoding novel
29	304.5	15.1	3732	23	AA592892	DNA encoding novel
30	304.5	15.1	3732	23	AA593885	DNA encoding novel
31	304.5	15.1	44668	21	AA81495	N. meningitidis pa
32	304.5	15.1	269223	22	AA28554	Genomic fragment #
33	292	14.5	607	24	AB24964	Oligonucleotide fo
34	292	14.5	607	24	AB24965	Oligonucleotide fo
35	286	14.2	1830121	17	AAT42063	Hemophilus influe
36	279.5	13.8	1086	20	AB090313	M. capsulatus gene
37	271	13.4	1005	20	AA212304	Neisseria meningit
38	271	13.4	1005	21	AA253632	Neisseria meningit
39	239.5	11.9	900	24	AB091605	M. capsulatus gene
40	220	10.9	607	24	AB024963	Oligonucleotide fo
41	220	10.9	607	24	AB024963	Oligonucleotide fo
42	219	10.8	569	24	AB037264	Oligonucleotide fo
43	219	10.8	569	24	AB037265	Oligonucleotide fo
44	217	10.7	306	24	AB75607	Human glycoprotein
45	188	9.3	1182	25	AB237910	N. gonorrhoeae nuc

ALIGNMENTS

RESULT 1
AAA37645
ID AAA37645 standard; DNA; 1239 BP.
XX
XX AAA37645;
AC
XX
XX 24-OCT-2000 (first entry)
DT
XX
XX N. meningitidis BASB05 coding sequence.
DE
XX
XX BASB05; diagnosis; microbial infection; invasive bacterial disease;
KW Neisseria meningitidis infection; upper respiratory tract infection;
KW bacteremia; meningitis; therapy; ss.
XX
XX Neisseria meningitidis.
OS
XX
FH Key Location/Qualifiers

FT	CDS		1..1239	/tag= a	BASB055
ET				/product= BASB055	
PV			WO200043517-A1.		
PD			27-JUL-2000.		
XX					
PP			19-JAN-2000; 200OMO-EP00425.		
PR			22-JAN-1999; 99GB-0001462.		
PR			29-JAN-1999; 99GB-0002069.		
PA	(SMIK)		SMITHKLINE BEECHAM BIOLOGICALS.		
XI					
PI	Thomnard J;				
DR	WPI; 2000-476199/41.				
PSDB; AAY90283.					
PT	Isolated BASB055 polypeptides, polymucleotides, and antibodies, the				
CC	polypeptides and polymucleotides are useful as vaccines for treating				
CC	and diagnosing a microbial infection such as a Neisseria meningitidis				
CC	infection _				
CS	Claim 7; Page 63; 82pp; English.				
XX	This sequence encodes the Neisseria meningitidis BASB055 polypeptide of				
CC	the invention. The BASB055 polypeptides and polymucleotides are useful				
CC	for diagnosing and treating microbial infections such as a Neisseria				
CC	meningitidis infection. They can also be used to treat any disease caused				
CC	by or related to infection by a bacteria, including upper respiratory				
CC	tract infection, invasive bacterial diseases (such as bacteraemia) and				
CC	meningitis.				
SQ	Sequence 1239 BP; 275 A; 332 C; 384 G; 248 T; 0 other;				
	Alignment Scores:				
	Pred. No.: 1,48e-157 Length: 1239				
	Score: 2019.00 Matches: 412				
	Percent Similarity: 100.00% Conservative: 0				
	Best Local Similarity: 100.00% Mismatches: 0				
	DQuery Match: 100.00% Indels: 0				
	Gaps: 21 Gaps: 0				
US-09-889-756A-2 (1-412)	x AAA37645 (1-1239)				
OY	1 MetAlapHeTyRAlApheylsaLwMcLaRGAlalALaleuaLAalaVAlaLeu	20			
Db	1 ATGGCTTTTAAgCtTTTTTAAGCGGAACGCCTGCCGTTCcCcGCAATG	60			
OY	21 ValleusSerSeCyaglYLvEglvgLYAspaALAagLnGLyGlInProAlAgLyArG	40			
Db	61 GTACTGTCTCTTGCGGTrAAAGCGGAGAAGCGCCGACCAGGGCGGAGCTGTGTGG	120			
OY	41 GlualAProAlAProValValGlyValValThrValHISPrOgLInTHrVALaleutnr	60			
Db	121 GAAGCCCCGCGCCCCTGCTCGGTgcGrAACCGTccATCCGAACCGTcgCATtgAcc	180			
OY	61 ValGIuleUProgLIARgleugJuseRLearqTRhalaspvalARGalaGINvalIGly	80			
Db	181 gfcGAGATTGCGGGGGCGTTGGMATGCTGCGTrAcCGCGATGTCGCGCCAAGTCGGC	240			
OY	81 GLylllellieglmYAArgLEUphegnIngUlgYSerTYValARGalagLYginPROleu	100			
Db	241 GGCAATCATCAAATAACCGCTGTCCAAAGAAGCACgtTAATGTCGTCCGCGACAAGCGCTG	300			
OY	101 TyrglnlIEASPserSerThrTYrgluAlasentLEUGlseRALargalaginLEuala	120			
Db	301 TRTCGARTCACAGttTCCActTRAteAGCAAAATCTGAAAAGCCGCGCGCAACTggCa	360			
OY	121 ThrAlaginalATHrlEuAlaySaIAspAIaspLEUALahRTyrILysProlEUAl	140			

Dd	361	ACGGCTCAGCGAAGCGCTTGCCAAAGCGGAGTGGCGATTGTCGGCATACAAAGCCTTGGTT	420
Qy	141	AlAlaagluAlaValSerArgIngluTyRAspAlaAlaValThrAlaIySAserSerAla	160
Dd	421	GCCCCCGAAGCCGTCACGCGCGAGGAATACAGTGTCTCGTAACGGCGAAACGTTCTGCC	480
Qy	161	GlunIaagIyValIySAIlaagInAlaAlaIleYsSerAlaIyIleasnLeuASArg	180
Dd	481	GAGGAGGTGTCAAAAGCAGCAGCGCGCAATCAAACTTGCAGCATTAATCTAGACCGT	540
Qy	181	SeArArgIleThrAlaProIleSergIyPheIleGIyInserIyValSergIuGIyThr	200
Dd	541	TGGGCAATTACCGCGCCGATTTCCGGCTTATATGCGTACGATCCAAAGTTTCGAAGTAGCG	600
Qy	201	LeuLeuASnaIaGIyASpThrThrValLeuAlaThrIleArgInThrAsnProMetTyr	220
Dd	601	CTGTTGAATGCGGCGATACAGCCGCTGTGGCAACCATCCGCGCAAAACCAATCCGATGAT	660
Qy	221	ValasnValThrGInSeraIaSerGIuValMetIySerIySergArgInIleAlaGIuGIy	240
Dd	661	GTGAACGTTACCCAGTCGTGCATCCGAAGTAGTAAGAAATGGCGCTCAGATAGCCGAAGGC	720
Qy	241	LysLeuLeuAlaAlaASpGIyValIleAlaValGIyIleIySPhaASpASpGIyThrValI	260
Dd	721	AAACTGCTGGCGGCGGATGTGTATTCGGTCGGCATCAAAATTGACAGCGACAGCTT	780
Qy	261	TyrProGIuIySGIyArGIeLeuPheAlaASpProValValasnIySergThrGIyGIIn	280
Dd	781	TACCTGTAAGAAAGCGCCGCTGCTGTTCGATCCGATCCGTCGTCAACGAATGACCGGTCAG	840
Qy	281	IleThrIeunArgAlaAlaValProASnaASpGIInasInIleIeunMetProGIyLeuTyVal	300
Dd	841	ATTACCCCTGCGCGCGCGCGGTACCGAAGCATCAGATATCTGATGCCGCTGTGTATGTG	900
Qy	301	ArgValIeunMetASpGIInAlaAlaValASpASnaIaPheValValProGIInAlaAlaVal	320
Dd	901	CGCGTCTGTATGACCAAGTGGCGGTGATACCCATTTGTGTGGCCGACAGCGCGCTA	960
Qy	321	ThrArgGIyAlaIyASpThrValMetIleValASnaIaGIInGIyMetGIuProArg	340
Dd	961	ACGCGCGGTGGGAAGATACCGGTATGTGTAATCCCAAGCGGATATGGAACCCCGC	1020
Qy	341	GIuValIThrValAlaGIInGIInGIyIThrASpTrpIleValIThrSergIyLeuIyASp	360
Dd	1021	GAGGTAAACGTTGGCAGACAGCAGGAGTACGAATGTGATTGTACGTGGGTCTGAAGGAC	1080
Qy	361	GIyASpIyValIaValaGIuGIyIleSeriIleAlaGIyIleThrGIyAlaIyASyValI	380
Dd	1081	GGGAGCAAGGTGTGTTCGAAGGATCAGTATGCCCGGTATACGGGTGCCAAAAAGTA	1140
Qy	381	ThrProIySGIuTyRAlaIaSerSergIuASnGIInAlaAlaIaProGIInSergIyValGIIn	400
Dd	1141	ACGCCCAAGATGGGGGTGTCTGAAGAACCAAGCCCGCGCTCATCCGGCTTCAG	1200
Qy	401	ThrIaIaSerGIuAlaIyThrIaIaSerGIuAlaGIu	412
Dd	1201	ACGGCATCTGAAGCCAAAACTGCTTCTGAAGCGGAA	1236
RESULT 2			
ABKS4080			
ID	ABKS4080 standard; DNA; 1239 BP.		
XX	ABKS4080;		
AC	ABKS4080;		
XX	05-JUN-2002 (first entry)		
DE	DNA encoding Neisseria cell surface polypeptide #2.		
XX	Cell surface protein; gene; ds; antibacterial; antimicrobial.		
KW	Neisseria meningitidis.		
OS	Neisseria meningitidis.		
XX	WO200216612-A2.		
PN			

PF 08-OCT-1999; 99W0-US23573.
 XX 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX (CHIR) CHIRON CORP.
 XX Frazer CM, Hickey E, Peterson J, Tettein H, Venter JC;
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlatto V;
 PI Rappoli R, Pizzi M;
 XX WPI: 2000-318079/27.
 DR
 XX
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N. gonorrhoea*.
 XX
 XX Claim 7, Page 622-628; 1760pp; English.

The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX Sequence 20352 BP; 5082 A; 5105 C; 5364 G; 4800 T; 1 other;

Alignment Scores:

Pred. No.: 1,32e-153 Length: 20352
 Score: 1989.00 Matches: 406
 Percent Similarity: 99.03% Conservative: 2
 Best Local Similarity: 98.54% Mismatches: 4
 Query Match: 98.51% Indels: 0
 DB: 21 Gaps: 0

US-09-889-756a-2 (1-412) x AAA81488 (1-20352)

QY 1 MetAlaPheTyrAlaPheLeuValMetArgAlaAlaLeuAlaAlaValAlaLeu 20
 Db 12358 ATGGCTTTTAAAGCTTTAAAGCGCATGCGCGCGCGGTGGCGCGCGGTGCATTG 12417
 QY 21 ValLeuSerSerCySgLyLyGlyGlyAspAlaAlaGlnGlyGlnProAlaGlyArg 40
 Db 12418 GTACTGTCGCTTCGCGGTAAAGCGGAGACCGCGGCGGCGGCGGCGGCTGCTCGG 12477
 QY 41 GluAlaProAlaProValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
 Db 12478 GAAAGCCCTCGCGCGCGTGTGCTGCTAAACCTCCATCCGCAACCGTGCATTGACC 12537
 QY 61 ValGluLeuProGlyArgGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
 Db 12538 GTCGAGTTGCGGGCGCTTGGAAATCGTGCATCCGCGGATGTCGCGCCCAAGTGGCG 12597

QY 81 GlyIleIleGlnLysArgLeuPheGlnGlySerTyrValArgAlaGlnProLeu 100
 Db 12598 GGCATCATCAAAAACCGCTGTTCCAAAGAGCAGTATATCGTGCAGACCGCTG 12657
 QY 101 TyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlnSerLysAlaArgAlaGlnLeuAla 120
 Db 12658 TATCAGATCAGACAGTTCCACTTATGAAGCAGGTCTGAAAGCGCGCGCAACCTGCA 12717
 QY 121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal 140
 Db 12718 ACGGCTCAGGCACACGCTTGCCTCAAGCGCATGCCGATTCGCGCATCAACCTTGTT 12777
 QY 141 AlaAlaGlnAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysSerAla 160
 Db 12778 GCGCGCGAAGCCGTCAGCGCGGAGAAATACGATGTCGGTAAACGCGGAACGTTCTGCC 12837
 QY 161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerLysAlaGlyIleAsnLeuAsnArg 180
 Db 12838 GAGCAGCGCTTAAAGCGCGCAGCGCATCAAAATCCCGCATCACCTGAACCGT 12897
 QY 181 SerArgIleThrAlaProLysSerGlyPheIleGlyGlnSerLysValSerGlnGlyThr 200
 Db 12898 TCGCGCATTTACCGCGCGCATTTCCGCTTATCGTCATCCAAAGTTCCGAAGTACG 12957
 QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
 Db 12958 TTGCTGAACCTGGCGCATGCGACCGTACTGGCAGCATCCGCCAACAACCATCCGATGTAT 13017
 QY 221 ValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlnGly 240
 Db 13018 GTGAACCTTACCAAGTGTGATCCGAAAGTATATAATGGCCGTCAGATAGCGAAGGC 13077
 QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Db 13078 AACTGTCGCGCGCGCATGTTGATTCGCGTCCGATCAAAATTTGACGACGACGATT 13137
 QY 261 TyrProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGln 280
 Db 13138 TACCTGAATAAAGCGCGCTGTTGGCGCATCCGCGCTCAACGAATGACCGGTGACG 13197
 QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnIleLeuMetProGlyLeuTyrVal 300
 Db 13198 ATTAACCTGGCGCGCGCGTACGAAACGATCAGAAATTTGATGATGCGGTCTGTATGTG 13257
 QY 301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
 Db 13258 CGCGTCTGATGACCAAGTGGCGGTGATACGCAATTTGTTGCGCGCAGACGCGGTA 13317
 QY 321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyLysMetGluProArg 340
 Db 13318 ACGGCGCGTCCGAAAGATACCGTATGATTTGTAATCCCAAGCGGTATAGCAACCGCGC 13377
 QY 341 GluValThrValAlaGlnGlnGlnGlyThrAsnThrIleValThrSerGlyLeuLysAsp 360
 Db 13378 GAGGTAAACGTTGCGCAACAGCAGGTACGAATTTGATTTGCTGCTGCTGGAAGAC 13437
 QY 361 GlyAspLysValValValGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysVal 380
 Db 13438 GGGGACAAGTGTGTGGGAAGCATCATGTCGCGGTATACCGGTGCGAAGAAAGTA 13497
 QY 381 ThrProLysGlnTyrAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
 Db 13498 ACGGCCAAGATGGGTGCTGTAACCAAGCCGCGCTCAATCCGCGGTTCAG 13557
 QY 401 ThrAlaSerGlnAlaLysThrAlaSerGlnAlaGln 412
 Db 13558 ACGGCATCTGAACCAACCTGCTTGAAGCGA 13593

RESULT 4
 AAF21611/c
 ID AAF21611 standard; DNA; 349980 BP.
 XX AAF21611;

QY 381 ThrProlysgluTTPAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
 Db 295242 ACCGCCAAGATGATGGCTGCTGTGAAAAACCAAGCCCGCGCTCAATCCGGGCTTACG 295183
 QY 401 ThrAlaSerGluAlaValThrAlaSerGluAlaGln 412
 Db 295182 ACGGCATCTGAAGCCAAACCTGCTTGTGAAGCGAAA 295147
 RESULT 5
 ID AAA81489 standard; DNA; 837096 BP.
 AC AAA81489;
 DT 04-DEC-2000 (first entry)
 XX N. meningitidis partial DNA sequence gmm_37 SEQ ID NO:37.
 DE N. meningitidis partial DNA sequence gmm_37 SEQ ID NO:37.
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KM Meningococcus B; MenB; de.
 XX Neisseria meningitidis.
 OS Neisseria meningitidis.
 PN MO200022430-A2.
 XX 20-APR-2000.
 PD 08-OCT-1999; 99WO-US23573.
 PF 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX (CHIR) CHIRON CORP.
 PA Frazer CM, Hickey E, Peterson J, Tettein H, Venter JC, Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R, Pizza M;
 PI WPI; 2000-318079/27.
 DR Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX Claim 7; Page 629-865; 1760pp; English.
 PS The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA8159 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have also been tried but none have successfully
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SQ Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
 Alignment Scores:
 Pred. No.: 1,19e-151 Length: 837096
 Score: 1989.00 Matches: 406
 Percent Similarity: 99.03% Conservative: 2
 Best Local Similarity: 98.54% Mismatches: 4
 Query Match: 98.51% Indels: 0
 DB: 21 Gaps: 0
 US-09-889-756a-2 (1-412) x AAA81489 (1-837096)
 QY 1 MetAlaPheTyrAlaPheTyrAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeu 20
 Db 361153 ATGGCTTTTATGCTTTTAAAGCGATGCGCGCGCGCTGGCTGCGCTGCAATTG 361094
 QY 21 ValLeuSerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40
 Db 361093 GTACTGTCGCTTTCGGGTAAAGCGGAGACGCGCGCGCGCGCGCTGCTGTCGG 361034
 QY 41 GluAlaProAlaProAlaValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
 Db 361033 GAAGCCCTTCGCGCGCGCTGCTGCGGTGCTGTAACCGTCCATCCGAAACCGTCCATTAC 360974
 QY 61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
 Db 360973 GTCGAGTTGCGGGGCGTTTGATGATCGTCGTAACCGCATGCGCGCGCGCGCGCGCG 360914
 QY 81 GlyTleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeu 100
 Db 360913 GGCATCATCCAAAACCCCTGTTCCAAAGAGGAGTATATCCGTCCGACAGCCCTG 360854
 QY 101 TyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAla 120
 Db 360853 TATCAGATTCACAGTTCACTTATGAAAGCAGGCTGGAAGCCGCGCGCGCAATGCGCA 360794
 QY 121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal 140
 Db 360793 ACGGCTCAGGCAACGCTTGGCAAGCGGATGCGGATTTGGCGGATTCGCAAGCTTGGTT 360734
 QY 141 AlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSerAla 160
 Db 360733 GCGGCCCAACCGCTCACCGCGAGATGATGATGATGATGATGATGATGATGATGATGAT 360674
 QY 161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleLeuLeuAsnArg 180
 Db 360673 GAGGAGGCGCTTAAAGCGGCGGCGGCGGCAATCAATCCCGCGCATCACCTGAACCT 360614
 QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGluGlyThr 200
 Db 360613 TCGGCGATTACCGCGCGGATTTCCGGCTTATCGGTACGTCACGTCGCAAGTTCCGAAGTACG 360554
 QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
 Db 360553 TTGGTGAACCTGCGGATGCGACCTGACGCACTGCGCAACCAATCCGATGTAT 360494
 QY 221 ValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGluGly 240
 Db 360493 GTGAACCTTACCCAGTGTGATCGAATGATGATGATGATGATGATGATGATGATGATGAT 360434
 QY 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Db 360433 AAACGTGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360374
 QY 261 TyrProGluLysGlyArgLeuLeuPheAlaAspProValValAlaGlnGluSerThrGlyGln 280
 Db 360373 TACCTGAAAAAAGCGCGCTGTTGCGCATCCGCGCGCTCAACGATGACGCGGTACG 360314
 QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuVal 300
 Db 360313 ATTACCTGCGCGCGCGCGATGCGAAGATGCGAAGATGCGAAGATGCGAAGATGCGAAG 360254
 QY 301 ArgValLeuMetAspGlnValAlaValAlaAspAsnAlaPheValValProGlnGlnAlaVal 320

Db 360253 CGCGTCTGATGAGCAACAGTGGCGGTGATACGATTTGTTGCGCCAGCAGCGGTA 360194
 QY 321 ThrArgGlyAlaIleAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Db 360193 ACGGGCGGTGGCAAGATACCGTATGATTTGATGATGCCAAGGGGATGAAACCCCGC 360134
 QY 341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
 Db 360133 GAGGTAAACGGTTGGCAACACAGGGTACGAATTGATTTACGTGGGTCTGAAGGAC 360074
 QY 361 GlyAspLysValIleValGluGlyIleSerIleAlaGlyIleThrGlyAlaLysLysVal 380
 Db 360073 GGGGACAAAGTGTGTGGAAAGCATCATCCCGGTATACGGGTGCAGAAAGGTA 360014
 QY 381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaIleAlaProGlnSerGlyValGln 400
 Db 360013 ACGCCCAAGATAGGCGCTCTGTGAACCAAGCGCGCCCTCAATCCGGCGTTGAG 359954
 QY 401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGln 412
 Db 359953 ACGGCATCTGAAGCCAAACCTGTTCTGAAGCGAA 359918
 RESULT 6
 ABZ39062
 ID ABZ39062 standard; DNA; 1236 BP.
 XX AC ABZ39062;
 XX 07-MAR-2003 (first entry)
 XX N. gonorrhoeae nucleotide sequence SEQ ID 2713.
 XX Antibacterial; infection; vaccine; gene therapy; gene; ds.
 XX Neisseria gonorrhoeae.
 XX W0200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizsa M, Maignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX P-PSDB; ABP78092.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX PT medicament for treating or preventing N. gonorrhoeae infection -
 XX PS Disclosure; Page 386; 815pp; English.
 XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
 XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
 XX CC antibodies that specifically bind to the proteins. The composition
 XX CC comprising the protein, nucleic acid or antibody is useful for the
 XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 XX CC infection, this may be in the form of a vaccine or gene therapy.
 XX CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
 XX CC molecules of the invention.
 XX SQ Sequence 1236 BP; 268 A; 334 C; 390 G; 244 T; 0 other;
 Alignment Scores:
 Pred. No.: 2.78e-151 Length: 1236
 Score: 1943.00 Matches: 394
 Percent Similarity: 97.57% Conserved: 8
 Best Local Similarity: 95.63% Mismatches: 10

Query Match: 96.24% Indels: 0
 DB: 25 Gaps: 0
 US-09-889-756a-2 (1-412) x ABZ39062 (1-1236)
 QY 1 MetAlaPheTyrAlaPheLysAlaMetArgAlaAlaAlaLeuAlaIleValAlaLeu 20
 Db 1 ATGGCTTTTATGCTTCTTAAGCGCATCGTCGCCCGCGTTCGCGCGCGTTCATTTG 60
 QY 21 ValLeuSerSerGlyLysGlyValAspAlaIleGlnGlyGlnProAlaGlyArg 40
 Db 61 GCACGTGCTTTCGCTTAAGAGCGAGAGCGGCGAGGGGGAGCTGCGAGTGG 120
 QY 41 GluAlaProAlaProValAlaGlyValIleThrValHisProGlnTrpValAlaLeuThr 60
 Db 121 GAACCCCGCGCGCGCTGCTGCGGTGTAACCTTCATCCGCAACCGTGCATTGACC 180
 QY 61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
 Db 181 GTCGAGTTGGCGGGCGCTTGGAAATCGTCGTCGCCCATGTCGCCCAAGTCGGC 240
 QY 81 GlyIleIleGlnLysArgLeuPheGlnGlySerTyrValArgAlaGlyGlnProLeu 100
 Db 241 GGCATCATCAAAACCGCTGTCGAAGAGCGATTATGTCGCGCGGCGAGCCGCTG 300
 QY 101 TyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAla 120
 Db 301 TATCAGATCAACACTTCACTTATGAAGCAGGTGTGAAGACGGCGCGCACTGGCA 360
 QY 121 ThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal 140
 Db 361 ACGGCACAGCAAGCTTGGCAAGCGGATGCGGATTGGCGGTTCAACCGCTGTT 420
 QY 141 AlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysSerAla 160
 Db 421 TCCGCCGATCCATCACTAAACAAAGTACGATGCTCGGTAAACGGGAAACGTTTCC 480
 QY 161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAspArg 180
 Db 481 GAGCGCGGCTTAAAGCGCGCGAGCGCGCATCAATCCCGCGCATCAATCAACCT 540
 QY 181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThr 200
 Db 541 TCGGCATTAACGGCGGATTTCCGGCTTATCGTCAGTCAAAAGTTCCGAAGTACG 600
 QY 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
 Db 601 CTGTGAATCGGGCGGATACACTGTTTATAGCCACATCCGCCAAACCAATCCGATGAT 660
 QY 221 ValAsnValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlnGly 240
 Db 661 GTGAACCTTACCAAGTCTGCATCCGAGTATGAATGCGCGCGCGAGATGCGAAGGC 720
 QY 241 LysLeuLeuAlaIleAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Db 721 AAGCTGTGCGCGGATGATGCGATTCGCGTGGCATCAATTTGACGACGATGCGTT 780
 QY 261 TyrTrpGlyLysGlyArgLeuLeuPheAlaAspProValValLeuGluSerThrGlyGln 280
 Db 781 TATCTGGAAGAAAGCGCGCTGCTGCTGATCCGACCTTGAGGATGACGCGGTACG 840
 QY 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
 Db 841 ATTAACCTTGCGCGCGCGATGATGAACGATCAAGATATCTGATGCGCGCGCTGTATGTG 900
 QY 301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
 Db 901 CGTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
 QY 321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Db 961 ACGCGCGGTGCGAAGATACCGTGAATGATGTAATGCCAGGCGCGGTATGAACCCCGC 1020


```

Qy 341 GluValThrValAlaGlnGlnGlnGlyThrSerThrIleValIleThrSerGlyLeuLysAsp 360
Db 1021 GAGGTAAAGCTGCGGCAACAGAGGATTCGATTGATTGATTCGCGTCTGAAGAC 1080
Qy 361 GlyAspLysValIleValGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysLysVal 380
Db 1081 GGGGACAGGCTGTTGGAGAGCATCATGCGCGGTATGACGGGGCGAAAAAGGTA 1140
Qy 381 ThrProLysGluThrPalaSerSerGluAsnGlnAlaIleAlaProGlnSerGlyValGln 400
Db 1141 ACGGCTAAAGATATGGCGCCCTCTGAAATATCAAGCTGCGCCCTCAAGCGGCTTCAG 1200
Qy 401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGln 412
Db 1201 ACGGATCTGAAGCCAACTGCTTCTGAAAGCAAA 1236

RESULT 7
AA211064
ID AA211064 standard; cDNA; 4879 BP.
AC AA211064;
XX
XX 01-NOV-1999 (first entry)
DE E. coli acra and acrb genes.
XX
XX Acra; acrb; tolC; organic solvent resistance; indigo production;
KM streptoid conversion; flooded crude oil treatment; ds.
XX
XX Escherichia coli.
XX
XX Key Location/Qualifiers
XX CDS 329..1522
XX FT /*tag= a
XX FT /product= acra
XX FT sig_peptide 329..400
XX FT /*tag= b
XX FT mat_peptide 401..1519
XX FT /*tag= c
XX FT CDS 1545..4694
XX FT /*tag= d
XX FT /product= acrb

JPI1221080-A.
XX
XX 17-AUG-1999.
XX
XX 09-FEB-1998; 98JP-0027537.
XX
XX 09-FEB-1998; 98JP-0027537.
XX
XX 09-FEB-1998; 98JP-0027537.
XX
XX (MEIJ ) MEIJ SEIKA KAISHA LTD.
XX
XX WPI; 1999-520718/44.
XX P-PSDB; AAY32854; AAY32855.
XX
XX Escherichia coli having organic solvent resistance - useful for
XX production of indigo, conversion of steroid and treating flooded
XX crude oil
XX
XX Claim 2; Page 6-8; 14pp; Japanese.
XX
XX This sequence encodes the acra and acrb proteins. The invention relates
XX to a method for the preparation of E coli having organic solvent
XX resistance or of increased organic solvent resistance including a step of
XX transforming the E. coli with at least one gene of acra, acrb and tolC.
XX The E. coli can be used for the production of indigo from indol,
XX conversion of a steroid such as cholesterol and treatment of flooded
XX crude oil.
XX
XX Sequence 4879 BP; 1157 A; 1215 C; 1343 G; 1164 T; 0 other;
Alignment Scores:

```

```

Pred. No.: 4,07e-53 Length: 4879
Score: 762.50 Matches: 168
Percent Similarity: 61.31% Conservative: 76
Best Local Similarity: 42.21% Mismatches: 129
Query Match: 37.77% Indels: 25
DB: 20 Gaps: 6

US-09-889-756a-2 (1-412) x AA211064 (1-4879)
Qy 11 AlaAlaIleuAlaIleAlaIleValAlaLeuValIleuSerSerCysGlyLysGlyAsp 30
Db 359 GCGGCTCTTGTAGTCTCTCTCAGGACAGCTTACCTTACAGATGT--GACGACAAACAG 415
Qy 31 AlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAlaProValIleGlyVal 50
Db 416 GCCCAACAGAGTGGCCAG-----CAGATGCCCGCGGTGGGTAGTA 457
Qy 51 ThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70
Db 458 ACAGTCAAAACTGAACTCTGCAAGATCATCAACGAGCTTCGGGTGACACAGTGTAC 517
Qy 71 ArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlu 90
Db 518 CGATGCCAGAGATTCCTCTCTCAAGTTAGCGGATTTCTGAAAGCTAATTCAGAA 577
Qy 91 GlySerThrValArgAlaGlyGlnProLeuThrGlnIleAspSerThrThrGlyAla 110
Db 578 GGTAGCGACATCGAAGAGAGTGTCTCTCTTACAGATTGATCTCGACCTATCAGGCG 637
Qy 111 AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAsp 130
Db 638 ACATACGACAGATCGAAAGGTGATCTGGCGAAAGCCAGGCTTCAGCCAAATATCGCGAA 697
Qy 131 AlaAspLeuAlaArgThrLysProLeuValAlaAlaGluAlaValSerArgGlnLys 150
Db 698 TTGACGCTGAATGTTTTCGAACATCTCTCGTACATCAGTACATGACAGAGTAC 757
Qy 151 AspAlaAlaValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAla 170
Db 758 GATGAGCTCTGGCTGTATGCGCAACAGCGAATGCTCGGTAACTGGCGGAAAGTGGCC 817
Qy 171 IleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 818 GTTGAAACTCGCGGAGTCAATCTGGCTTACACCAAGTCACTCTCGATTAGGGTGGC 877
Qy 191 IleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValIleu 210
Db 878 ATTGTAAGTCGAACGTACGGAAGGCGCATTTGTCAGAACGCTCAGCGACTGGCCTG 937
Qy 211 AlaThrIleArgGlnThrAsnProMetThrValAsnValThrGlnSerAlaSerGluVal 230
Db 938 GCAACGCTGACGACATCTGATCCGATCTGATGATGACCCAGTCCACACACGACTTC 997
Qy 231 MetLysLeuArgArgGlnIleAlaGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
Db 998 CTGGCCTGAAACAGGAACGTGCGAATGCGACCTGAAACAGACGCGCAAGCCAA 1057
Qy 251 ValGlyLysPheAspArgGlyThrValThrProGlyLysGlyArgLeuLeuPheAla 270
Db 1058 GTGTCACTGATCACACAGTACGCGCATTAAGTTCGCGAGACGCTACGCTGAATTCCT 1117
Qy 271 AspProValValAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
Db 1118 GAGCTTACGTTATATACGACACAGCTGGTCTATACCTTACGCGCTATCTCCGAACCG 1177
Qy 291 GlnAsnIleLeuMetProGlyLysThrValArgValLeuMetAspGlnValAlaValAsp 310
Db 1178 GATCACACTCTGCTGCGGATATGTTCTGTCGCGACGCTGAGAGAGGCTTAATCA 1237
Qy 311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrVal 328
Db 1238 AACGTAATTTAGTCCGCAACAGGGCGTAACCGGTACGCCGCTGCGCATGCCACCGTA 1297

```


Qy 329 MetIleValaAsnAlaGlnGlyGlyMetCjUpProArGluValThrValAlaGlnGln 348
 Db 1298 CTGTAGTGGCGGATGACAAAGTGAACCCGTCGATCGTCGACAGCCAGCTATT 1357
 Qy 349 GlyThrAsnTrpIleValThrSerGlyLeuAspGlyAspIleValValGlnGly 368
 Db 1358 GGCATTAAGTGGCGGATGACAAAGTGAACCCGTCGATCGTCGATTAAGTGG 1417
 Qy 369 IleSerIleAlaGlyIleThrGlyAlaIleValIleValThrPro 382
 Db 1418 CTG-----CAGAAAGTGGCTCTGATGATTAACGACGACCGCA 1453
 Qy 383 LysGluTrpAlaSerSerGlnAsnGlnAlaAlaProGlnSerGlyValGln 400
 Db 1454 CAAGAAGTACCGCTGATTAATTAACGACGACCGCA-----ACGGGTCTCAG 1501

RESULT 8
 ID ABS78861 standard; DNA; 7888 BP.
 XX ABS78861;
 AC
 XX 17-DEC-2002 (first entry)
 DT
 XX E. coli CFT073 genomic sequence #28.
 DE
 XX Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
 KM urinary tract infection; open reading frame; ORF; uropathogenic;
 KM antibacterial; atropathic; nephrotropic; gene; ds.
 XX
 OS Escherichia coli.
 XX WO200259320-A2.
 PN
 XX 01-AUG-2002.
 PD
 XX 19-OCT-2001; 2001WO-US46833.
 PF
 XX 19-OCT-2000; 2000US-242412P.
 PR
 XX (MISC) WISCONSIN ALUMNI RES FOUND.
 PA
 XX Blatner FR, Welch RA, Burland VD;
 PI
 XX WPI; 2002-691532/74.
 DR
 XX
 PT New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
 PT useful for preventing or treating E. coli CFT073 infection in humans or
 PT livestock
 XX
 PS Claim 1; Page 120-124; 765pp; English.

CC The present invention relates to polynucleotide sequences from the
 CC genome of the pathogenic Escherichia coli strain CFT073. Almost all
 CC the sequences present in E. coli CFT073 are absent in the previously
 CC sequenced laboratory strain K-12. The polynucleotide sequences of
 CC the invention are useful for preventing, diagnosing or treating
 CC E. coli CFT073 infection in humans or livestock. The polynucleotide
 CC sequences are useful for preventing urinary tract infections and
 CC pyelonephritis. Likewise, the polypeptides encoded by the different
 CC open reading frames (ORF1-5) are useful for generating a vaccine
 CC against uropathogenic E. coli strains. ABS78834-ABS79085 represent
 CC genomic sequences from E. coli strain CFT073.
 CC
 XX
 SQ Sequence 7888 BP; 1884 A; 1854 C; 2150 G; 2000 T; 0 other;

Alignment Scores:
 Pred. No.: 7.89e-48 Length: 7888
 Score: 701.50 Matches: 162
 Percent Similarity: 60.63% Conserves: 69
 Best Local Similarity: 42.52% Mismatches: 133
 Query Match: 34.74% Indels: 17
 DB: 24 Gaps: 7

US-09-889-756a-2 (1-412) x ABS78861 (1-7888)
 Qy 7 IysAlaMetArgAlaAlaIleValAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 1064 AAAATATATGAAGTATATAGCAACATCTGTAGTGCAATGCTGCTTATCGGTTGT--- 1120
 Qy 27 LysGlyValAspAlaAlaGlnGlyGlnProAlaGlyValArgIleValProAlaProVal 46
 Db 1121 -----GATATATAGCAAAAGTAACATTCATCCCGTCAGAAACAGAA----- 1162
 Qy 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 Db 1163 GTCGGCTTTCATCGCTCAATCTCAACCCGTTTCGGTATCTCATGTAATTAACGGACGT 1222
 Qy 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnIleValArg 86
 Db 1223 ACCAGTCTCGCTCGCTGAAAGTACGCCGAGTTGGGGAATTATTCAGAAACGC 1282
 Qy 87 LeuPheGlnGluGlySerTyrrValArgAlaGlyGlnProLeuTyrrGlnIleAspSerSer 106
 Db 1283 TTATTTAAGGAAGGTATCTGTCTAAGCTGACAGCTCTCTACCGATTTAGCCGCC 1342
 Qy 107 ThrTyrrGluAlaAsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeu 126
 Db 1343 AGTTATCAGGCTCATGAAATGAAGCCCGGACAGCATTTACACAGACAGCA---CTG 1399
 Qy 127 AlaIleValAspAlaAspLeuAla---ArgTyrrLeuProLeuValAlaAlaGlnAlaVal 145
 Db 1400 GTAAAGCCGATTCAGCAAGAAAGCGCTTAATCCGACTGTGGTGAAGAAAGAAACGGTGT 1459
 Qy 146 SerArgGlnGluTyrrAspAlaAlaValThrAlaIleValSerAlaGlnAlaGlyValIle 165
 Db 1460 TCAACAAGAGATGCTATGATGCTCAAGCTCACTGACCAACAAATAAAGCCAGTGAAG 1519
 Qy 166 AlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAla 185
 Db 1520 GCGAAAAAGCCCACTGAAACCTGGCGCGCATTAATCTTACCTGACACGATTAACCGCA 1579
 Qy 186 ProIleSerGlyPheIleGlyGlnSerIleValSerGlnGlyThrLeuLeuAlaGly 205
 Db 1580 CCGATTTGGGGGCTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1639
 Qy 206 AspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrrValaAsnValThnGln 225
 Db 1640 CAGGATACACCTTACGACTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1699
 Qy 226 SerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeuAlaAla 245
 Db 1700 TCCAGTGTGATTTATTAGTGTGCGTAACAG-----TCACGTGGGACCAAC 1747
 Qy 246 AspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrrProGluTyrrGly 265
 Db 1748 AGTGACACCATGCGCTTCATCTTATTCGAAAGTGGCAACACTACAGCGAAAAAGGG 1807
 Qy 266 ArgLeuLeuPheAlaAspProValValaGlnGluSerThrGlyGlnIleThrLeuArgAla 285
 Db 1808 CGCTGGAACCTACCGCAAGTCGGGTGATGAATCTTACCGCTTCCGGAATTAACGGCA 1867
 Qy 286 AlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrrValaArgValLeuMetAsp 305
 Db 1868 ATTTTCCCAATCCACACAGCAAGTATTAACGGGAATGTTTGTTCGGCTGCTGAT 1927
 Qy 306 GlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLys 325
 Db 1928 GAAGGGGTATGTAAACGCTATTCTCGCGCTTCAACAGGCGCTTACCGCGATGCTTAA 1987
 Qy 326 Asp-----ThrValMetIleValaAsnAlaGlnGlyGlyMetCjUpProArGluValThr 343
 Db 1988 GGCATTAAGTGGCGGATGACAAAGTGAACCCGTCGATCGTCGATTAAGTGG 2047
 Qy 344 ValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLys 363

Db 2048 ACGGAGAAAGCTATGCTGATTAATGCTGCTGTAACGCGCTGCACCAAGCGGACCGA 2107
 QY 364 ValValValGluGlyIleSerIleAlaGlyIleThrGlyAlaIleValIleThrProlys 383
 Db 2108 CTGATTTGTTGAGGT-----TCTGCCAAAGTCACTTCAAGCGCAGACCGCTCAAGCTTT 2161
 QY 384 Glu 384
 Db 2162 GAA 2164
 RESULT 9
 ID AAF28550 standard; DNA; 99629 BP.
 XX AAF28550;
 AC AAF28550;
 XX 04-APR-2001 (first entry)
 DT
 XX Genomic fragment #37.
 DE
 XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KM bronchopulmonary; endocarditis; meningitis; ss.
 XX
 OS Moraxella catarrhalis.
 XX MO200078968-A2.
 XX 28-DEC-2000.
 XX 16-JUN-2000; 2000MO-US16649.
 XX 18-JUN-1999; 99US-0140121.
 XX PA (INCYTE) INCYTE GENOMICS INC.
 PT Lagace RE, Patterson C, Berg KL;
 DR WPI; 2001-041427/05.
 XX
 PT Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 PS Claim 1; Page 391-415; 545pp; English.
 XX
 CC The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX
 XX Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;
 Alignment Scores:
 Pred. No.: 1.02e-40 Length: 99629
 Score: 631.50 Matches: 167
 Percent Similarity: 50.97% Conservative: 70
 Best Local Similarity: 35.91% Mismatches: 139
 Query Match: 31.28% Gaps: 89
 DB: 22 Indels: 10
 US-09-889-756a-2 (1-412) x AAF28550 (1-99629)
 QY 21 ValIleuSerSerGlyIleGlyIleAlaAlaGlnGlyIleProAlaGlyArg 40
 Db 1311 GTGCTGACGGCTGCTCAACAGACGATTCACAG-----GTGCGACCA 1355

QY 41 GluAlaProAlaProValValGlyValValThrValHisProGlnThrValAlaLeuThr 60
 Db 1356 CAATGCGCAATCCGACGATGATGTTCAAACTTATGCTCCGAACTGGCCAGTATT 1415
 QY 61 ValGluLeuProGlyAlaGlyLeuGluSerLeuArgThrAlaAspValAlaGlnGlnValGly 80
 Db 1416 CAAAGTTTTCAGAGCGGTGTGCGCATTGAAGCTTCAAGAGGTCCCGCCAGGTACG 1475
 QY 81 GlyIleIleGlnIleValArgLeuPheGlnGlnGlyIleSerTyValArgAlaGlnProLeu 100
 Db 1476 GCGATTATGATGACAGACTGCTTCTATGAGAGTATGATGTCGACGACGACGACCTTTA 1535
 QY 101 TyrglnIleAspSerSerThrTy----- 108
 Db 1536 TATCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1595
 QY 109 -----GluAlaAsnLeuLysSerAlaArgAlaGlnLeu 119
 Db 1596 AACGCTGAGCGCGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1655
 QY 120 AlaThrAlaGlnAlaThrLeuAlaIleValAlaAspAlaAspLeuAlaArgTyIleProLeu 139
 Db 1656 GCGCTCAAGAGCGCAATTTGGCTCAGCGCAGCGCAGATGATGATGATGATGATGATGATGAT 1715
 QY 140 ValAlaAlaGlnAlaValSerArgGlnGluTyValAspAlaAlaValThrAlaValArgSer 159
 Db 1716 TTGGAGGTGAGCGATCTCTCGTCAAGCTTTGACCAACGATTAACCGCATGCGCAACC 1775
 QY 160 AlaGlnAlaGlyValIleValAlaAlaGlnAlaAla----- 170
 Db 1776 GCTGAAGCGAATGTTTGGCTGCGCAGCGCGCAGTTCGCGACGACGATGATTAATTTCAC 1835
 QY 171 -----IleIleSerAlaGlyIleAsnLeuAsnArgSerArg----- 182
 Db 1836 AGTGCAACAGCATCCGTCGCAAGTGCACGCGCCCAATCTAAACGCAACCGCATGATGCA 1895
 QY 183 -----IleThrAlaProIleSerGlyPheIleGlyIleSerIleValSerGlu 198
 Db 1896 GGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1955
 QY 199 GlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsnPro 218
 Db 1956 GGTGACTGTGATCCGCAATTCACCAACCAATGCTGATCAATTAATCGTTGATGAA 2015
 QY 219 MetTyValAsnValThrGlnSerAlaSerGluValMetIleValArgGlnIleAla 238
 Db 2016 GTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2075
 QY 239 GluGlyIleLeuLeuAlaAlaAspGlyValIleAlaValGlyIleIleValPheAspAspGly 258
 Db 2076 TCGGTCAAGTA-----GGTCAAGCGCGTGCAGATGCTCAAGTGGTGTGAAGATGCT 2129
 QY 259 ThrValTyProGluIleValArgLeuLeuPheAlaAspProValValAsnGluSerThr 278
 Db 2130 CAAACTTATCAATCTTGAAGCAATTAATGCTTCCCAATCAGCAAGGATGATGATGATGATGAT 2189
 QY 279 GlyIleIleThrLeuArgAlaAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeu 298
 Db 2190 GGCACATGAT 2249
 QY 299 TyrValArgValIleuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGln 318
 Db 2250 TTTGTTATGCTGCTTAATTCAAAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2309
 QY 319 AlaValThrArgGlyAlaIleVal-----AspThrValMetIleValAlaGlnGlnGly 336
 Db 2310 GCGATCATGCTGATCACCAAAAGCGCAAAACCAAGTTTATGCTTAACGACGATTAACAA 2369
 QY 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyIleThrAsnProIleValThrSer 356
 Db 2370 ATCGAGTTTGCACGCGGATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2429
 QY 357 GlyLeuIleAspGlyAspIleValValValGlnGlyIleSerIleAlaGlyIleThrGly 376

RESULT 13
AAH84642
ID AAH84642 standard; DNA; 1395 BP.
XX
AC AAH84642;
XX
DT 26-SEP-2001 (first entry)
XX
DE E. coli growth and proliferation related coding sequence SEQ ID NO:270.
XX
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
KM bacterial infection; microorganism; ds.
XX
OS Escherichia coli.
XX
PN MO200134810-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000MO-US30950.
XX
PR 09-NOV-1999; 99US-0164415.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen K, Zyskind J;
XX
DR WPI: 2001-335933/35.
XX
PT P-PSDB; AAG98971.
XX
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
for screening for homologous genes and for designing expression vectors
XX
PS Claim 9; Page 337-339; 522pp; English.
XX
CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
growth and proliferation related proteins given in AAG99078 and AAG98830
to AAG98999. (1) can be used as potential targets for the generation of
new antimicrobial agents, and for identification of compounds which
interact with the gene products of (1). In addition the expression of
(1) and the purification of the proteins, the purified proteins can be
used to generate reagents and screen small molecule libraries or other
candidate compound libraries for compounds that can be further developed
to yield novel antimicrobial compounds. In addition, nucleic acid probes
complementary to (1) that are specific for particular species of
microorganisms can be used to identify particular microorganism species
in clinical specimens, therefore, providing a rapid and dependable
method by which to identify the causative agents of a bacterial
infection. Also, antibodies generated against proteins translated from
mRNA transcribed from proliferation-related sequences can also be used
to screen for specific microorganisms that produce such proteins in a
species-specific manner. AAH84371 and AAH84670 represent sequences in a
primers used in the isolation of E. coli growth and proliferation
related sequence, which are used in an example from the present
invention.
XX
SQ Sequence 1395 BP; 360 A; 372 C; 383 G; 280 T; 0 other:

Alignment Scores:
Pred. No.: 8.77e-27 Length: 1395
Score: 435.50 Matches: 124
Percent Similarity: 47.94% Conservative: 62
Best Local Similarity: 31.96% Mismatches: 176
Query Match: 21.57% Indels: 26
DB: 22 Gaps: 7

US-09-889-756A-2 (1-412) x AAH84642 (1-1395)
QY 22 LeuSerSerCysGlyLysGlyGlyAspAlaIaIaGlnGlyGly-GlnProAlaGly----- 39
DB 240 CTCCTCCGAGTGCACCCCGAGGGCGAGCAAAACAAAGCGAGCAATCGCCAGCGGTGTGCG 299

QY 40 -----ArgGluAlaProAlaProValGlyValValThrValHisProGlnTh 56
DB 300 ACGTGTATGCGTTCCGCCCATTTAGCCCGGTTACAGCGCGGACCGCCGTGAACAAGGC 359
QY 56 rValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValAr 76
DB 360 AGTTCCCGCTTACTCCCGGCTTGGCACCATTTACCGCTTAATCGCTTAATCGCTTAC 419
QY 76 gAlaGlnValGlyGlyLeuIleGlnLysArgLeuPheGlnGlnGlySerThrValArgAl 96
DB 420 CAGCGCGGTGAGCGCCCACTGATAGCGTTACATTTCCAGAAAGGCCAGAGTCAAGC 479
QY 96 aGlyGlnProLeuThrGlnIleAspSerSerThrThrGlyAlaAsnLeuIleSerAlaAr 116
DB 480 AGCGATTTACTGCGCAAAATTAAGCCACGACGATTCAAAGTTGACATTAACAAGCCCA 539
QY 116 gAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaAlaAspAlaAspLeuAlaArgTy 136
DB 540 GGGCAACTGGCAAAAGATTAAGCCACGCTTGCCCAAGCCCGCGCTGACCTGGCGCTTA 599
QY 136 rLysProLeuValAlaAlaGluAlaValSerArgGlnGlyThrAspAlaAlaValThrAl 156
DB 600 TCAACAACTGGCAAAACCAATCTCGTTCCGCGAGAGCTGATGCCAAGCGGCT 659
QY 156 aLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyI 176
DB 660 GGTCAAGTGAACCAAGCAAGCCACATTAAGCTGATGAACCAAGGTTCCAGCGGCACT 719
QY 176 eAsnLeuAsnArgSerArgGlyLeuThrAlaProIleSerGlyPheIleGlyGlnSerLysVa 196
DB 720 GCAACTCGATCGAGCGGATTTACCGCACCGATGCGGTGGGTCTCAAGCGGT 779
QY 196 lSerGluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnTh 216
DB 780 TGATGTTGTAACCAATCTCCAGTGTGATTAACACCGGATGCTGTGATACCCAGAC 839
QY 216 rAsnProMetThrValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgI 236
DB 840 GCATCTCATGATTTAGTTCCTTACCGCCGGAAGGATATGCTCATCGTAAGTCAGGC 899
QY 236 nIleAlaGluGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyLe--LysPh 255
DB 900 GCAGAAACCGGAAACCGCTGTGTGAGAACCTGGATCGCAACCACTCGAAGAAATT 959
QY 255 eAspAspGlyThrValThrProGluLysGlyArgLeuPheAlaAspProValValAs 275
DB 960 AAGTGAAGGACG-----CTGTTAAGTCTTACATTAACCAATCGA 998
QY 275 nGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMe 295
DB 999 TGGCAGTACCGGTAGCATTTAAAGTGAAGCACCGCTTAATATATCAGATATGCGCTT 1058
QY 295 tProGlyLeuThrVal-----ArgValLeuMetAspGlnValAlaValAspAsnAlaPh 313
DB 1059 TCCCAATCACTTTGTTAAACCGCGCATGTATGTCGAC-----ACCGAACAACCGCGT 1112
QY 313 eValValProGlnGlnAlaValThrArgGlyAlaLys---AspThrValMetIleValAs 332
DB 1113 AGTGATCCCAACACCGCCCTGCAATGGCAATGAGGCAATTTGTCTGGGTGCTGAA 1172
QY 332 nAlaGlnGlyLysMetCysProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTr 352
DB 1173 TAGGAAAACAAAGTGCAGAAACATCTGCGAGACCGCGGATTCAGAGACGTCAAGAAAGT 1232
QY 352 pIleValThrSerGlyLeuLysAspGlyAspLysValValValGluGlyLe----- 369
DB 1233 GGTGATCCGTGACGATTTCTGCGGCGATTCGCGTGTACAGAGCGGCAATTCATCGCTT 1292
QY 370 -----SerIleAlaGlyIleThrGlyAlaLysLeuValThrProLysGluTr 385
DB 1293 GACCGAAGGGGCAAAAGTGAAGTGTGAAGGCCAGAGCCCACTTATCTCGGAAGAGAA 1352

XX Nucleotide sequence of PA14 mexa and mexB.
DE Human pathogen; virulence polypeptide; virulence factor;
KW pathogenic infection; Pseudomonas aeruginosa infection; ss.
XX Pseudomonas aeruginosa.
OS
PN WC0927129-A1.
XX
PD 03-JUN-1999.
XX
PF 25-NOV-1998; 98WO-US25247.
XX
PR 25-NOV-1997; 97US-0066517.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
PI Rahme LG, Tan M, Tsongalis J;
XX
DR WPI; 1999-357851/30.
XX
PT Virulence factors useful in developing disease treatments
PS Disclosure; Fig 14B; 228pp; English.
XX
CC The present sequence represents a Pseudomonas aeruginosa nucleic acid
CC sequence. P. aeruginosa is an opportunistic human pathogen present in
CC soil water and plants. The specification describes virulence polypeptides
CC and nucleic acid sequence encoding such polypeptides. These sequences
CC can be used to identify a compound which is capable of decreasing the
CC expression of a pathogenic virulence factor. Compounds that inhibit
CC the expression or activity of virulence factor polypeptides can be
CC used to treat pathogenic infections, especially where the infection
CC is a P. aeruginosa infection.
CC note: the sequences given in the specification were poorly legible, and
CC in some instances assumptions were made as to the identity of the base;
CC it is therefore possible that the sequence given below is not entirely
CC correct.
XX
SQ Sequence 1008 BP; 205 A; 331 C; 320 G; 152 T; 0 other;
XX
XX
Alignment Scores:
Pred. No.: 1,37e-21 Length: 1008
Score: 370.50 Matches: 80
Percent Similarity: 58.29% Conservative: 43
Best Local Similarity: 37.91% Mismatches: 73
Query Match: 18.35% Indels: 15
DB: 20 Gaps: 5
US-09-889-756a-2 (1-412) x AAX98220 (1-1008)
QY 205 GlyAspThrThrValLeuAlaThrIleArgInThrAsnProMetTyValAsnValThr 224
DB 1 GGCAGGCAAAACGGATGCGACCGCTGCACAGCTGCACCGCATCTAGTGAAGTCTCACC 60
QY 225 GluSerIleAspGluValMetLysLeuArgArgInIleAlaGluGlyLysLeu---Leu 243
DB 61 CAGCGCTCCACCGCGCTGTTGCGCATGCGCGCAACGCGCACGCGGCTGAGAGCGC 120
QY 244 AlaIleAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyProGlu 263
DB 121 GCCCGGCAACACGCTGCGAAGGTCCTCTGAAGCTGGAGGCGGAGGCAATACCGCTG 180
QY 264 LysGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGlnIleThrLeu 283
DB 181 GAAGCGCGCTCGAATCTCCGAGGTTCTCCGACGAGCAACGCGCTCGTCCATC 240
QY 284 ArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyValArgValLeu 303
DB 241 CGCGCGCTGTTCCCAACCCGCAACAGAGCTGCTGCCCGCATGTTGTTCAACGCGCAG 300

QY 304 MetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGly 323
DB 301 TTGAGAGAAAGCGCTCAAGCAAGACCATCTCGTCCGACGAAAGCGTGAACCCGCGAC 360
QY 324 AlaIleAsp-----ThrValMetIleValAlaAsnAlaGlnIleGlyMetGluProArgGlu 341
DB 361 CTCAGAGGCCAGGCTACCGCGCTGCTGTGAACCGCGCAACAAAGCTGAGCTCGCGGTG 420
QY 342 ValThrValAlaGlnGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuLysAspGly 361
DB 421 ATCAAGGCCACCGCGGATCGCGACAGTGTGCTCAACGAGCTGAAACCGCGGC 480
QY 362 AspLysValValValGluGlyIleSerIleAla-----GlyIleThrGlyAlaLysLys 379
DB 481 GACAAAGATCATTAACGAGCGCTGCAATGCTGTCAGCCGCGGTGCGAG---GTCAAGACC 537
QY 380 ValThrProLysGluTrpAlaSerSerGluAsnGlnAlaIleAlaProGlnSerGlyVal 399
DB 538 GTGCCGCGCAAGATGTGCGGTCCGCGCAAGCGCGACCGCTCGTCCG----- 585
QY 400 GlnThrAlaSerGluAlaLysThrAlaSerGlu 410
DB 586 -----GCCAAACCGACAGCAAG 603
RESULT 16
AAS79919
ID AAS79919 standard; cDNA; 2112 BP.
XX
AC AAS79919;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #15723.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG15732.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 1; SEQ ID No 15723; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2112 BP; 547 A; 601 C; 532 G; 432 T; 0 other;

Alignment Scores:

Pred. No.:	1,34e-18	Length:	2112
Score:	339.00	Matches:	105
Percent Similarity:	46.56%	Conservative:	78
Best Local Similarity:	26.72%	Mismatches:	134
Query Match:	16.79%	Indels:	77
		Gaps:	11

US-09-889-756A-2 (1-412) X AAS81865 (1-2112)

QY 6 PheylsAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCys 25
 Db 947 TTCGAGAACTGCCCTGACCGCATATCACCCCTGATCATATATCCACACAGTG 1006
 QY 26 GlyylsGlyArgAlaAlaGlnGlyGlyGlnProAlaGlyArgGlnAlaPro----- 43
 Db 1007 -----ACTTACCTCGGACGACGAGATGCCACCGCA 1039
 QY 44 -----AlaProValValGlyValValThrValHis-----ProGlnThrVal--- 57
 Db 1040 GCCGGAACACCATGACCGTACCGGTACGTCAGTTCAGTCACTGCTTCG 1099
 QY 58 -----AlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArgThrAla 73
 Db 1100 GTTATGTACCCCGCTATGACCATGACCGTTT-GGTACAGGTGACGCGGTCCACCGCAG 1158
 QY 74 AspyAlaArgAlaGlnValGlyGlyLeuLeuGlnArgLeuPheGlnGlySerTyr 93
 Db 1159 GAA-----GGGCTGAGTCAGTACCGC-----GCATAC 1185
 QY 94 ValArgAlaGlyGlnProLeuTyrGlnLeuSerSerThrTyrGlnAlaAsnLeuGlu 113
 Db 1186 GACAGCGGTGACGATTAATGCGGTGAAGACACGCGGCCATGAACCGGTATGAA 1245
 QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaSerAla 131
 Db 1246 TACAACCGCCGCGACATACGACATGCAAGTGTCTTCAAAAGCCCGACGTGACGC 1305
 QY 132 -----AspLeuAlaArgTyrTyrProLeuValAlaAlaGlnAlaValSer 146
 Db 1306 AATATCGCGCAATTGACGCTAATCGTATCGAAGAACTGCTGCTACTCACTACATCAGT 1365
 QY 147 ArgGlnGlyTyrAspAlaAlaValThrAlaValSerArgSerArgAlaGlyValValSerAla 166
 Db 1366 AAGCAAGAGTACGATGAGCTCTGCGTATGCGCAACGCGGATGCTGCGGTATGCT 1425
 QY 167 AlaGlnAlaAlaLeuSerSerArgGlyLeuLeuLeuSerArgGlyLeuThrAlaPro 186
 Db 1426 GCGAAGCTGCGCTTGAAGTGCACAGATCAATCTGTTACACCAAG-----GCACCT 1479
 QY 187 IleSerGlyPheIleGlyGlnSerTyrValSerGlyGlyThrLeuLeuAlaGlyAsp 206
 Db 1480 CTCGATTAGCGGTGCGATTCGTGACGAGGAGCGCATTTGTACAGAGGCTCAG 1539

QY 207 ThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAlaValThrGlnSer 226
 Db 1540 GCGATCGCTGCGCAACCGTGACAGAACTTGATCGTATGATGACCGCACTCC 1599
 QY 227 AlaSerGluValMetLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAsp 246
 Db 1600 AGCAACACACTCTTCCGCTGAAACAGAACTGCGCAATGCGACGCTGAAACAGAGAAC 1659
 QY 247 GlyValIleAlaValGlyIleLysPheAspAspGlyThrValTyrProGlyLysGlyArg 266
 Db 1660 TGCAAACCCAAAGTGTACTGATCACCAGTACCGCTTAACTTCCGCGACAGCGTACG 1719
 QY 267 LeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAla 286
 Db 1720 CTGGAATTTCTGACGTTACCGTTCATGACAGACCTGCGTATACCTTACCGGCTATC 1779
 QY 287 ValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGln 306
 Db 1780 TTCCCGAACCAGATCACACTCTGCTCCGCGC----- 1812
 QY 307 ValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp 326
 Db 1812 ----- 1812
 QY 327 ThrValMetIleValAsnAlaGlnGlyGlyMetGluPro-----ArgGluValThr 343
 Db 1813 AATATCAGCAATGAGATATGACTGCGCGGCTTGGACTTCAATCACCCGATCCCACTG 1872
 QY 344 ValAlaGlnGlnGlnGlyThrAsnThrIleValThrSerGlyLeuLysAsp----- 360
 Db 1873 ATCATTAAGTATGATGCGCACCCGTTGGTGTCAAAACATGATGACGAAACCAATACT 1932
 QY 361 GlyAspLysValValAlaGlnGlyIleSerIleAlaGly 373
 Db 1933 GCGGCGACGACCTGATTCGCGGAAGTACGCTTTCAGA 1971
 Db
 RESULT 18
 AAS82449
 ID AAS82449 standard; cDNA; 2310 BP.
 XX
 AC AAS82449;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #18253.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYS-E-) HYSBQ INC.
 XX
 PI Dermanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG18262.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 18253; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 2310 BP; 562 A; 600 C; 634 G; 514 T; 0 other;

Alignment Scores:

Pred. No.:	2,58e-17	Length:	2310
Score:	324.00	Matches:	86
Percent Similarity:	48.62%	Conservative:	37
Best Local Similarity:	33.9%	Mismatches:	105
Query Match:	16.05%	Indels:	25
DB:	23	Gaps:	2

US-09-889-756A-2 (1-412) x AAS82449 (1-2310)

QY 22 LeuSerSerCysGlyLysGlyValAspAlaIleGlnGly-GlnProAlaGly----- 39
 DB 1527 CTCCCGAGTGCACCCAGCGGCGAGCAACAGCGAGCATTCGACGCGGTGCTCG 1586
 QY 40 -----ArgLunAlaProAlaProValAlaGlyValValHisProGln 56
 DB 1587 ACGTGTATGCGTCCGCGCCATTAGCCCGGTTCAGCGCGGCGAGCGCGGTGCAAGGC 1646
 QY 56 rValAlaLeuThrValAlaGluLeuProGlyArgLeuGluSerLeuArgThrAlaSerValAr 76
 DB 1647 AGTCCCGCTTACTCCACCGCGCTTGCGACCATTAACCGCCCTAATACGTTAGCGTGGC 1706
 QY 76 gAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAl 96
 DB 1707 CAGCGCGTGGAGCGGCACTGATGAGCTTATTCATTTCCAGGAAGCGCAGCGTCAAGC 1766
 QY 96 aglyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGluSerAlaAr 116
 DB 1767 AGGGATTTACTGGCAAAATATGACCCCGACGCACTTAAAGTTGATTCAGCAAGGCCA 1826
 QY 116 gAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTyr 136
 DB 1827 GGGCACTCGCAAAATAAGCAACGCTTGCCCAAGCGCGCTAATCTGGCGCTTAA 1886
 QY 136 rLysProLeuValAlaAlaGlnAlaValSerArgGlnGlnLysrAspAlaValThrAl 156
 DB 1887 TCACAACACTGGCAAAACCAATCTGTTCCCGCAGAGGTGATGCCAAGCGCGCT 1946
 QY 156 aLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyT 176
 DB 1947 GGTGATGAACACGAGGACCACTTAAGGCTGATGAAGCAAGCGCTTCCAGCGCGAGCT 2006
 QY 176 eAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysVa 196
 DB 2007 GCAACTGACTGAGCGCGGATTAACGCAACGATGAGTGCCTTGTCTCAAGCAGGT 2066
 QY 196 lSerGlnGlyThrLeuLeuAlaGlyAspThrThrValLeuAlaThrIleArgGlnTh 216

DB 2067 TGATGTTGTTACCAAACTTCAGTGTGATATCCACCGGATCGTGTATCACCAGAC 2126
 QY 216 rAsnProMetTyrValAlaValThrGlnSerAlaSerGlyValMetLysLeu----- 233
 DB 2127 GCATCTATGATATTATCTTATACCTCGCGGAAGGATATCGCTACCGTAGGCAAGC 2186
 QY 234 -----ArgArgGlnI 237
 DB 2187 GCAGAAAGCGGAAACCGCTGTGTGAGAGCCTGGAGTCGACCAACTGAAAGAAATT 2246
 QY 237 eAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIle 249
 DB 2247 AAGTGAAGGACCGCTGTTAAGTTAAGTAAACCAATA 2283

RESULT 19

ID ABZ40695 standard; DNA, 1176 BP.
 AC ABZ40695;
 XX 07-MAR-2003 (first entry)
 DT 07-MAR-2003 (first entry)
 DE N. gonorrhoeae nucleotide sequence SEQ ID 5979.
 XX Antibacterial; infection; vaccine; gene therapy; gene; ds.
 XX Neisseria gonorrhoeae.
 OS Neisseria gonorrhoeae.
 XX W0200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002MO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 PR (CHIR-) CHIRON SPA.
 XX Fontana MR, Piazza M, Maignani V, Monaci E;
 PI WPI, 2003-058415/05.
 DR P-PSDB; ABP79725.
 DR New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 XX Disclosure; Page 624; 815pp; English.
 PS The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
 CC molecules of the invention.

SO Sequence 1176 BP; 301 A; 275 C; 379 G; 221 T; 0 other;

Alignment Scores:

Pred. No.:	9.22e-17	Length:	1176
Score:	313.00	Matches:	116
Percent Similarity:	43.57%	Conservative:	67
Best Local Similarity:	27.62%	Mismatches:	176
Query Match:	15.50%	Indels:	62
DB:	25	Gaps:	11

US-09-889-756A-2 (1-412) x ABZ40695 (1-1176)

QY 7 LysAlaMetArgAlaAlaAlaAlaAlaAlaAlaValAlaLeuValLeuSerSerCysGly 26
 DB 7 AAATGATGAAGGCGCGCTGTTGGCGCGGTGCGCGCGCA----- 48

```

Qy 27 LysGlyIysAspAlaIleGlnGlyGlyProAlaGlyArgGlyAlaProAlaProVal 46
Db -----GGCGTTTGGGGCGGAGTGTCTTATCTGAAGCCGAGCCGAGCTGCT 96
Qy 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArg 66
Db 7ATATTACGAAACGCTAGCGCGGCGGCGATACAGCCGCGAGGTTTCCGCGACGGCGAG 156
Qy 67 LeuGlySerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnIysArg 86
Db 157 ATTTCGGCGTCAACCTGTATCGCTATCGTCCGCGGAGGCTTGCGGCGACATTAAGAGCTT 216
Qy 87 LeuPheGlnGlySerThrValArgAlaGlyIleProLeuThrGlnIleAspSerSer 106
Db 217 TATGTCAAACTCGGCGAACAGGTCAAAAAGGCGGATTTGATTCGCGAATCAATTCGACC 276
Qy 107 ThrTyr-----GlnAlaAsnLeuGlySerAlaArgAlaGlnLeu 119
Db 277 ACGGACACCAACACGATCATATGGAATAATCCAAATTGGAACGATATCAGCGCAAGCTG 336
Qy 120 AlaThrAlaGlnAlaThrLeuAlaIysAlaAspAlaAspLeuAlaArgTyrLeuProLeu 139
Db 337 GTGTCCGACACGATTCGATTCGGCAGCGCGGAGAAAGAAATTAAGCTCAGCGCGCTTG 396
Qy 140 ValAlaAlaGlnAlaValSerArgGlnGlyIleThrAspAlaAlaValThrAlaLysArgSer 159
Db 397 TGAAGGATGATGCGACCTTAAGAGAGATTTGGAACCGCGCAGATGCTTGCCGCC 456
Qy 160 AlaGlnAlaGlyValIysAlaAlaGlnAlaAlaIleIysSerAlaGlyIleAsnLeuAsn 179
Db 457 GCCAAAGCCAAATGTTGCGAGTTGAAGCTTTAATACACAGACCAAAATTTCCATCAT 516
Qy 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
Db 517 ACCGCCAGATGCGATTTGGGCTACACGCGCATTCACGCGACGATGAGACGCGACGTGTG 576
Qy 193 GlnSerIysValSerGlyIleThrLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
Db 577 GCGATTCCTCGTGAAGAGGCGCAGACTGTGAACCGCGCG----- 615
Qy 213 IleArgGlnThrAspPrometTyrValAsnValThrGlnSerAlaSerGlyValMetLys 232
Db 616 -----CAGCTACGCGCAGCATTTGCAATTCGCGAAT-----CTGAGATATATGTG 663
Qy 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
Db 664 AACAAATGCAGATTCGCGAGGCGGATATTCACAAAGGAGGAGCGCGCGCAGATATTCG 723
Qy 251 ValGlyIleLysPheAspAspGlyThrValTyrProGlnIysGlyArgLeuLeuPheAla 270
Db 724 TTTCAGATTTTGTCCGAACCGGATACG-----CCGATTAGCGCAAGCTCGACAGCTC 777
Qy 271 AspProValVal-----AsnGlySerThrGlnGlnIleThr 282
Db 778 GACCCCGGCTGACACGATGTCCTCGGCGGCTACCAACACAGATACGATACGCGCTTC 837
Qy 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
Db 838 AATCGCGCTATTTATATGCTTCCTGTTGTCGCGAATCCGCGACGCGCAACCTGCCACG 897
Qy 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValPro 316
Db 898 GGGATGACGACGAAATACGCTGGAATACGACGCTGGAATAATGTTGCTTATTCG 957
Qy 317 GlnGlnAlaVal---ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
Db 958 TCGCTGACCGTGAATAATCGCGCGGCAAGCGCTTCGACCGCTGTGGGCGGCGACGCG 1017
Qy 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrIleValThr 355
Db 1018 AAGGACGTGGAACCGCAATTCGCGGATGTAAGACAGTATGAATACGGAAGTGA 1077
Qy 356 SerGlyLeuIysAspGlyAspLysValValValGlnGlyIleSerIleAlaGlyIleThr 375

```

```

Db 1078 AGCGGTTGAAAGGGGGGCAAGTGTGATCTCCGAATATACCGCGCGA-GCAGCA 1136
Qy 376 GAlaValLysLysValThrProLysGlyIleThrAlaSerSerGlnIleAlaAlaPro 395
Db 1137 GGAAGCGGGAACGCGCCTTAGG-----CGGCCCGCGCGCCG 1175

RESULT 20
AAZ12306
ID AAZ12306 standard; DNA; 1179 BP.
XX
AC AAZ12306;
XX
XX 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae complete ORF85 sequence.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
XX
OS Neisseria gonorrhoeae.
XX
XX
XX W09924578-A2.
XX
XX
XX 20-MAY-1999.
XX
XX
XX 09-OCT-1998; 98MO-IB01665.
XX
XX 01-SEP-1998; 98GB-0019016.
XX 06-NOV-1997; 97GB-0023516.
XX 14-NOV-1997; 97GB-0024190.
XX 18-NOV-1997; 97GB-0024386.
XX 27-NOV-1997; 97GB-0025158.
XX 10-DEC-1997; 97GB-0026147.
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
PI WPI: 1999-337407/27.
XX
XX DR P-PSDB; AAY38884.
XX
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
XX
XX Claim 9; Page 430; 524pp; English.
XX
XX Nucleotide sequences AAZ11972-Z12358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX
XX
XX Sequence 1179 BP; 304 A; 275 C; 378 G; 222 T; 0 other;
SO

Alignment Scores:
Pred. No.: 1.64e-16 Length: 1179
Score: 310.00 Matches: 115
Percent Similarity: 43.57% Conservative: 68
Best Local Similarity: 27.38% Mismatches: 176
Query Match: 15.35% Indels: 62
DB: 20 Gaps: 11

US-09-889-756a-2 (1-412) x AAZ12306 (1-1179)
Qy 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
Db 7 AAATGATGAATAAGCGGCTGTGCGCGCTCGCGCGCAACGCTT----- 54

```

```

QY 27 LysGlyValAspAlaAlaGlnGlyGlnProAlaGlyArgGluAlaProAlaProVal 46
Db 55 -----TGGGGCGGATGGTCTTATCTGAAGCCGGAACCCGAGCTGCT 96
QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
Db 97 TATATTACGAACCGATGAGCGCGCGATATACAGCCGAGGTTTCCGCAACGGCGCAG 156
QY 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuGlnLysArg 86
Db 157 ATTTCGCGCTCAACCTGTATCGTTCGCGCGAGCTTCGGGGCAGATTAAAAAGCTT 216
QY 87 LeuPheGlnGluGlySerThrValArgAlaGlyGlnProLeuThrGlnLysAspSerSer 106
Db 217 TATGTCAAACTCGGCGCAACAGTCAAAAAGGCGCATTTGCGGAAATCAATTCGACC 276
QY 107 ThrTyr-----GluAlaLeuGluSerAlaArgAlaGlnLeu 119
Db 277 ACGCAGACCAACAGATGATATGAAAATTCAAATTTGAAACGCTATCAGCGCAAGCTG 336
QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLysProLeu 139
Db 337 GTGTGCGACAGATTGATTTGGCAGCGCGAAGAAATTAAGCTCAGCGCGCTTG 396
QY 140 ValAlaAlaGluAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSer 159
Db 397 TGGAAAGATGATCGACCTCTAAAGAAAGATTGGAAGCCGCGAGATGCGCTTGCGGCC 456
QY 160 AlaGluAlaGlyValLysAlaAlaGlnAlaAlaLeuLysSerAlaGlyLeuAsnLeuAsn 179
Db 457 GCCAAAGCAATGTTGCCGAGTGAAGGCTTTAATTCAGACAGCAAAATTTCCATCAT 516
QY 180 Arg-----SerArgLeuThrAlaProLysSerGlyPheLeuGly 192
Db 517 ACCGCCAGTGGATTTGGGCTACACGGCGATTACCGGAGCATGAGCGGACCGGTG 576
QY 193 GlnSerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
Db 577 GCGATTCCTCCGGAAGAGGGCGACACTGTGAACCGCGG----- 615
QY 213 IleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232
Db 616 -----AGTCTACGCGACGATTCCTCAATTTGGCGAAT-----CTGATATATGTTG 663
QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
Db 664 AACAAATGCAATTCGCGAGCGCGATATTAACCAAGTGAAAGCGGCGCAGATATTTCG 723
QY 251 ValGlyLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAla 270
Db 724 TTATGATTTTGTTCGAAACCGGATACG-----CCGATTAAAGCGAAGCTCGACAGGCTC 777
QY 271 AspProValVal-----AsnGluSerThrGlyGlnIleThr 282
Db 778 GACCCCGGCTGACCAAGATGCTCGGGCGGCTACCAACAGCATACGATACGCGCTTC 837
QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
Db 838 AATGCGCTATATTATTTGCCCGTTCCTTGTTCGCAATTCGAGCGCAAACTGCCACG 897
QY 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
Db 898 GGGATGACGACGCAATACGGTTGCAATCGAGTGAATAATGTTGCTTATTCG 957
QY 317 GlnGlnAlaVal---ThrArgGlyValLysAspThrValMetIleValAsnAlaGlnGly 335
Db 958 TCGCTGACCGTGAATAATTCGCGGCGGAGCGCTTCGATACGCGTGGGCGGAGCGC 1017
QY 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsnTyrPileValThr 355
Db 1018 AAGGCACTGAAACCGCAATTCGAGCGGATATGAAACAGTATACCGAAGTGA 1077
QY 356 SerGlyLeuLysAspGlyAspLysValValGlnGlyLysSerIleAlaGlyIleThr 375

```

```

Db 1078 AGCGGTTGAAAGAGGGGGGACAAAGTGTCTATCTCCGAAATTAACCGCGCGCA-GCAGCA 1136
QY 376 GlyAlaLysLysValThrProLysGluThrPalaSerSerGluWanGlnAlaAlaPro 395
Db 1137 GGAAGGCGCGAAGCGCGCCCTAGG-----CGGCGCGCGCGCG 1175

RESULT 21
ID AAF21607 standard; DNA; 349980 BP.
AC AAF21607;
XX 13-MAR-2001 (first entry)
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX KM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX ds.
XX OS Neisseria meningitidis.
XX MO20006791-A1.
XX 09-NOV-2000.
XX 08-MAR-2000; 2000MO-US05928.
XX 30-APR-1999; 99US-0132068.
XX 08-OCT-1999; 99MO-US23573.
XX 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMI.C RES.
XX P1 Piza M, Hickey E, Peterson J, Tettein H, Venter JC, Masignani V,
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
PI Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
XX DR
XX PT Neisseria meningitidis B full length genome sequence and open reading
XX frames are used to detect, treat and prevent Neisserial infections -
XX Claim 7; Appendix A; 692pp; English.
XX PS
XX CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;
Alignment Scores:

```

```

Pred. No.: 1.77e-13 Length: 349980
Score: 309.50 Matches: 119
Percent Similarity: 43.69% Conservative: 68
Best Local Similarity: 27.80% Mismatches: 186
Query Match: 15.33% Indels: 56
DB: 21 Gaps: 11
US-09-889-756a-2 (1-412) x AAF21607 (1-349980)

Qy 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
Db 272470 AAAATGATGAATGGCGCGGTGCGCGGTCCGCGGCGCA----- 272511
Qy 27 LysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlyAlaProAlaProVal 46
Db 272512 -----GGCGTTGGGGCGGATGCTTATCTGAAAGCCCGAGCCGCGCT 272559
Qy 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
Db 272560 TATATTACGGAAACGGTCAGCGCGCGACATCAGCCGAGCTTCTGCAACAGGGGAG 272619
Qy 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnLysArg 86
Db 272820 ATTTCCGCGTCAACCTGCTATCGTCCGCGCGAGCAGCAGTGAAGATTAAGTACTT 272679
Qy 87 LeuPheGlnGlnGlySerTyValArgAlaGlyGlnProLeuTyArgIleAspSer--- 105
Db 272680 TATCTCAACCTCGGGCAACAGGTTAAAGCCGATTTGATGGCGAAATCAATTCAGCC 272739
Qy 106 -----SerThrTyGlnLysLeuLeuLeuGlnSerAlaArgAlaGlnLeu 119
Db 272740 TCCGAGACCAATACGCTCAATACGGAATAATCCAAAGTTGGAACGATACAGCGAGCTG 272799
Qy 120 AlaThrAlaGlnAlaThrLeuAlaAlaAlaAspAlaAspLeuAlaArgTyLysProLeu 139
Db 272800 GTGTCGACACGATTGATTGGCGACGCGGAGAAAGAAATTAAGCGCAGCGCGCTTA 272859
Qy 140 ValAlaAlaGlnAlaValaValaValaValaValaValaValaValaValaValaVala 159
Db 272860 TCGAAGAGAAACGCGACTTCCAAAGAGATTGGAAACGCGCAGATGCTTGGCCGCC 272919
Qy 160 AlaGlnAlaGlyValaValaValaAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
Db 272920 GCCAAAGCATGTTGGCGAGCTGAAGCTTAATCAGACAGCAAAATTTCCATCAAT 272979
Qy 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
Db 272980 ACCGCCGAGTCGAATTGGGCTACACGCGCATTCACCGACAGATGAGCGACGCTGGTG 273039
Qy 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
Db 273040 GCGATTCCTCGTGAAGAGGGGCGACAGCTGTGAACCGCGG----- 273078
Qy 213 IleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerGluValaMetLys 232
Db 273079 -----CAGTCTACGCGCAGCATTTGTCATTAATGCGAAT-----CTGATATATGATTG 273126
Qy 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValaIleAla 250
Db 273127 AACAAATGCGAGATTGCGAGAGCGGATTTACCAAGGAGGAGCGGGCGAGATTTTTCG 273186
Qy 251 ValGlyIleLysPheAspAspGlyThrValTyProGlnLysGlyArgLeuLeuPheAla 270
Db 273187 TTTACGATTTTGTCCGAACCGGATACG-----CCGATTAAGCGCAACCTGACAGCGCTC 273240
Qy 271 AspProValVal-----AsnGlnSerThrGlyGlnIleThr 282
Db 273241 GACCCCGCGGTGACACAGATGTCGCGCGGCTTACAAACACAGACAGATACGCGCTTCC 273300
Qy 283 -----LeuArgAlaAlaValaProAsnAspGlnAsnIleLeuMetPro 296
Db 273301 AATGCGGTACTATATATATGATGCGGTTGCTGTCGATTCGCGAGCGGCAACCTGCGCACG 273360

```

```

Qy 297 GlyLeuTyValArgValLeuMetAspGlnValaAlaValaAspAsnAlaPheValaPro 316
Db 273361 GGGATGACGACGAGAAATACGTTGAATGACGCGCTGAAATAATGCTGATTAATCCG 273420
Qy 317 GlnGlnAlaVala---ThrArgGlyAlaLysAspThrValaMetIleValaAlaGlnGly 335
Db 273421 TCGCTGACCGTGAATAATCCGCGCGGAGCGCTTGTGTGCGGTGGTGGCGAGCGC 273480
Qy 336 GlyMetGluProArgGlnValaThrValaAlaGlnGlnGlnGlyThrAsnThrIleValaThr 355
Db 273481 AAGCGCGGAGAACCGAAATCCGACCGGATAGACACGATGAAATACCGAGTAA 273540
Qy 356 SerGlyLeuLysAspGlyAspLysValaValaGlnGlyIleSerIleAlaGlyIleThr 375
Db 273541 AGCGGTTGAAGAGAGGGGACAAAGTGCTCATCTCCGAATTAACCGCGCGCA-GCAACA 273599
Qy 376 GlyAlaLysLysValaThrProLysGluTrpAlaSerSerGluLeuGlnAlaAlaAla--- 394
Db 273600 GGAAGCGGAGAACCGCGCTAGCGCGCGCGCGCGCATTAACGAATATGCGCTCTGA 273659
Qy 395 ProGlnSerGlyValaGlnThrAla 402
Db 273660 ACAACGAAACGTTTCAGACGCA 273683

RESULT 22
AAAB1490
ID AAAB1490 standard; DNA; 1437668 BP.
XX
AC AAAB1490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN W0200022430-A2.
XX
PD 20-APR-2000.
XX
PE 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scariato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414
CC represent specifically claimed Neisseria meningitidis genome DNA
CC sequences; AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAAB1254 to
CC AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to
CC AAAB1452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies

```

CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B: against all serotypes;
 CC and/or against all pathogenic Neisseriae: identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX Sequence 1437668 BP: 344338 A: 353206 C: 385074 G: 355045 T: 5 other;

Alignment Scores:

Pred. No.:	9.81e-13	Length:	1437668
Score:	309.50	Matches:	119
Percent Similarity:	43.69%	Conservative:	68
Best Local Similarity:	27.80%	Mismatches:	186
Query Match:	15.33%	Indels:	56
DB:	21	Gaps:	11

US-09-889-756a-2 (1-412) x AA81490 (1-1437668)

QY 7 LysAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 572470 AAAATGATGAATATGGCGCGGTTCGCGCGCGCGCA----- 572511
 QY 27 LysGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlyAlaProAlaProVal 46
 Db 572512 -----GGCGTTGGCGCGCGATGCTTATCTGAAGCCCGAGCCGAGCT 572559
 QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 Db 572560 TATATTACGGAACGCTCAGCGCGCGCAATCCAGCCGCGTTCGCAACAGGGAG 572619
 QY 67 LeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuGlnLysArg 86
 Db 572620 ATTTCCGCGTCAACCTGATTCGTCGCGCGCGCAATCGCGGAGATTAAATATCTT 572679
 QY 87 LeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnLeuSer--- 105
 Db 572680 TATGTCAATCTCGGCAACAGTTAAAGGCGATTGATTCGCAAAATCAATTCCACC 572739
 QY 106 -----SerThrTyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeu 119
 Db 572740 TCGCAGACCAATACGCTCATACGGAATAATCCAAATTGGAACGATACGCGCAACTG 572799
 QY 120 AlaThrAlaGlnAlaThrLeuAlaAlaAspAlaAspLeuAlaArgTyrLysProLeu 139
 Db 572800 GTGTCGCGACAGATTGATTCGCGCGCGAGGAAGAAATATACCTTCAGCGCGCTTA 572859
 QY 140 ValAlaAlaGlnAlaValSerArgGlnGluTyrAspAlaAlaValThrAlaLysArgSer 159
 Db 572860 TCGAAGAAAACCGCACTTCCAAAGAGATTGGAAAAGCGGAGAGATGGCTTGGCGGC 572919
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLeuSerAlaGlyLeuAsnLeuAsn 179
 Db 572920 GCCAAAGCCAAATGTTGCCGAGCTGAAGCTTATACAGACAGCAAAATTTCCATCAAT 572979
 QY 180 Arg-----SerArgLLeuThrAlaProGlyPheLeuGly 192
 Db 572980 ACCGCCAGTTCGGAATGGCTACACCGCATTAACCCCAACGATGACGCGAGCTGGTG 573039
 QY 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 573040 GCGATTCTCGTGAAGAGGGGACACTGTGAACCGCGC----- 573078
 QY 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232

Db 573079 -----CAGTCTACCGCGAGATTGTCCAATTGGCGAAT-----CTGATATGATGTTG 573126
 QY 233 LeuArgArgGlnLeuAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 573127 AACAAATGCAAGATTCGCGAGGGGATTTACCAAGGTGAAGGGCGGCGAGATTTTCG 573186
 QY 251 ValGlyLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPheAla 270
 Db 573187 TTTACGATTTTGTTCGCAACCGGATACG-----CCGATTAGCGGAAGCTGACAGGCTC 573240
 QY 271 AspProValVal-----AsnGlnSerThrGlyGlnLeuThr 282
 Db 573241 GACCCCGCGTGCACACAGATGTCGCGCGGCTTACCAACAGCATACGATACGCGCTTC 573300
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 Db 573301 AATGCGGTCTACTATTATGTCCTGTTGTGCGCAATCCGACGGCAAACTGCGCAGC 573360
 QY 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
 Db 573361 GCGATGACAGCAGCAATACGTTGAATCGACGCGTGAATAATGCTGATTTATTCG 573420
 QY 317 GlnGlnAlaVal---ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 573421 TCGCTGACCGTGAAAAATCGCGCGCGCAAGCGCTTGTGCGCGTGGGTGCGACGCG 573480
 QY 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnThrPileValThr 355
 Db 573481 AAGCGCGCGAAGCGCAAAATCCGACCGGTATGAGACAGTATGATACCGAAGTAAA 573540
 QY 356 SerGlyLeuLysAspGlyAspLysValValAlaGlnGlyLysSerIleAlaGlyLeuThr 375
 Db 573541 AGCGGTTGAAGAAGGGGACCAAGTGTGATCTCCGAATTAACCGCCCGCA-GCAACA 573599
 QY 376 GlyAlaLysLysValThrProLysGluTyrPheLysSerSerGluAsnGlnAlaAla--- 394
 Db 573600 GGAAGCGCGCAACGCGCCTTAGCGCGCGCGCGCGCGATTAACGAATATGCGCTGA 573659
 QY 395 ProGlnSerGlyValGlnThrAla 402
 Db 573660 ACAAGGAACGTTTCAGACGCA 573683

RESULT 23
 AAS97298
 ID AAS97298 standard; DNA; 1176 BP.
 AC AAS97298;
 XX 12-MAR-2002 (first entry)
 DE Neisseria meningitidis virulence gene #103.
 XX
 KM Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
 XX infection; Gram-negative bacteria; antimicrobial; ds.
 OS Neisseria meningitidis.
 XX
 PN WO200185772-A2.
 PD 15-NOV-2001.
 XX
 PF 08-MAY-2001; 2001MO-GB02003.
 PR 08-MAY-2000; 2000GB-0011108.
 PA (MICR-) MICROSCIENCE LTD.
 XX
 PI Tang C;
 XX
 DR WPI; 2002-066593/09.
 XX P-PSDB; AA073013.
 XX

PT New peptide encoded by operon including virulence genes of *Neisseria meningitidis*, useful as vaccine component for treating or preventing meningitis and for identifying antimicrobial drug

PS Claim 1; Page 416-418; 423pp; English.

XX The invention relates to a peptide (I) encoded by an operon (II) of *Neisseria meningitidis* including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in a Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by *Neisseria* or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAS97196-AAS97305 represent *N. meningitidis* virulence genes and related PCR primers of the invention.

XX Sequence 1176 BP; 304 A; 276 C; 379 G; 217 T; 0 other;

Alignment Scores:

Pred. No.:	2,39e-16	Length:	1176
Score:	308.00	Matches:	114
Percent Similarity:	43.57%	Conservative:	69
Best Local Similarity:	27.14%	Mismatches:	176
Query Match:	15.26%	Indels:	62
DB:	24	Gaps:	11

US-09-889-756a-2 (1-412) x AAS97298 (1-1176)

QY 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 DB 7 AAATGATGAATGCGCGCTGTTGCGCGCGCGCGCGCA-----48
 QY 27 LysGlyArgAspAlaAlaGlnGlyGlyProAlaGlyArgGlyAlaProAlaProVal 46
 DB 49 -----GCGGTTGGGCGGATGCTTATCTGAAGCCCGAGCCCGAGGCTGCT 96
 QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 DB 97 TATATTACGGAACCGGTAGCGCGCGCGCAATCAACCGCGAGTTCGCAACAGGGAG 156
 QY 67 LeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArg 86
 DB 157 ATTTCGCGCTCAACCTGATCGGTGCGCGCGCGCATGCGGCAATTAAGAACTT 216
 QY 87 LeuPheGlnGlnGlySerThrValArgAlaGlyGlnProLeuThrGlnIleAspSer--- 105
 DB 217 TATGTCAAACTCGGCGCAACAGTTTAAAGGCGATTTGCGGAATCAATTCGACC 276
 QY 106 -----SerThrTyrgLulAlaLeuGlnLeuGlnSerAlaArgAlaGlnLeu 119
 DB 277 TCGCAGACCAATAGCCTCAATACGAAATTCCAATTTGAAACGTATCAGCGAGCTG 336
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrlsProLeu 139
 DB 337 GTTCGCGCAGATTCATTCGCGCGCGCGAGGAAGAAATTAAGCCGTCAGCGCGCTTG 396
 QY 140 ValAlaAlaGlnAlaValSerArgGlnLysAlaAlaValThrAlaLysArgSer 159
 DB 397 TGAAGAGTATGCGACCGCTTAAGAGATTTGAAACGCGACAGATGCTTCGCCGC 456
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 DB 457 GCCAAGCCATGTTGCCGAGCTGAGGCTCTAATCAGACAGCAAAATTTCCATCAT 516
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 DB 517 ACCGCCGAGTGGGAATGGCTACACGCGCATTCACGACAGATGAGCGAGGTGGTG 576
 QY 193 GlnSerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 DB 577 GCGATTCTCGTGAAGGGGCGACAGCTGTGAACCGCGCG-----615

QY 213 IleArgGlnThrAspProMetTyrValAsnValThrGlnSerAlaSerGlyValMetLys 232
 DB 616 -----CAGTCTACGCCGAGCATTTGCTCAATTTGGCGAAT-----CTGATATGATGTTG 663
 QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 DB 664 AACAAATGACAGATTTGCGAGCGGATATTAACAAATGAAAGCGCGCGAGATATTTCC 723
 QY 251 ValGlyLysPheAspArgGlyThrValTyrProGlnLysGlyArgLeuLeuPheAla 270
 DB 724 TTACGATTTTTCGCCGAACCGGATACG-----CCGATTAAAGCGAAGCTGACAGCGTTC 777
 QY 271 AspProValVal-----AsnGlnSerThrArgGlnIleThr 282
 DB 778 GACCCCGCGGCTACACAGATGTGTCGGGGCGGTACAAACAGCATGATAGCGGCTTCC 837
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 DB 838 AATCGGCTCTACTATTATGCCCCGTTGTCGCCGAATCCGACGCAAACTCGCCACG 897
 QY 297 GlyLeuThrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValPro 316
 DB 898 GCGATGACGACGCAATACGTTGAATGACGCGTGTGAATAATGCTGATTTATTCG 957
 QY 317 GlnGlnAlaVal---ThrArgLysAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 DB 958 TCGCTGACCGTGAATAATTCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1017
 QY 336 GlyMetGluProArgGlyValThrValAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 355
 DB 1018 AAGCGCGCGGAACCGGAATTCGCGACCGGTATGAGACAGTATGAAATCGAAGTAAA 1077
 QY 356 SerGlyLeuLysAspGlyAspLysValValGlnGlyLysSerIleAlaGlyIleThr 375
 DB 1078 AGCGGTTGAAGAGGGGCGCAAGTGTGATCTCCGAATTAACGCCGCCGCGA-GCGCA 1136
 QY 376 GlyAlaLysLysValThrProLysGlnLysLysSerGlnGlnGlnAlaAlaPro 395
 DB 1137 GAAAGCGCGGAACCGCGCTTAGG-----CGGCGCGCGCGCGCG 1175

RESULT 24
 AAS97248
 ID AAS97248 standard; DNA; 1177 BP.
 AC AAS97248;
 XX 12-MAR-2002 (first entry)
 DT
 DE
 DE *Neisseria meningitidis* virulence gene #53.
 KW Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
 KM infection; Gram-negative bacteria; antimicrobial; ds.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO200185772-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 08-MAY-2001; 2001WO-GB02003.
 XX
 PR 08-MAY-2000; 2000GB-0011108.
 XX
 PA (MICR-) MICROSCIENCE LTD.
 XX
 PI Tang C;
 XX
 DR WPI; 2002-066593/09.
 DR P-PSDB; AAU72963.
 XX
 PT New peptide encoded by operon including virulence genes of *Neisseria meningitidis*, useful as vaccine component for treating or preventing

PT meningitis and for identifying antimicrobial drug
 XX
 PS Claim 1; Page 224-226; 423p; English.

CC The invention relates to a peptide (I) encoded by an operon (II) of
 CC *Neisseria meningitidis* including virulence genes, or a related molecule
 CC having a 40% sequence similarity at the peptide or nucleotide level in a
 CC Gram-negative bacterium, or its functional fragment, for therapeutic or
 CC diagnostic use. (I) and (II) are useful in the manufacture of a
 CC medicament for treating or preventing a condition (e.g., meningitis)
 CC associated with infection by *Neisseria* or Gram-negative bacteria. The
 CC product is useful for veterinary treatment and in a screening assay for
 CC the identification of an antimicrobial drug. The vaccines have
 CC prophylactic applications. AAS97196-AAS97305 represent *N. meningitidis*
 CC virulence genes and related PCR primers of the invention.

XX Sequence 1177 BP; 304 A; 276 C; 379 G; 218 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	1177
Score:	308.00	Matches: 114
Percent Similarity:	43.57%	Conservative: 69
Best Local Similarity:	27.14%	Mismatches: 176
Query Match:	15.26%	Indels: 62
DB:	24	Gaps: 11

US-09-889-756A-2 (1-412) x AAS97248 (1-1177)

QY 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 7 AAATATGTGAATGCGGCTGTGGCGGCGCGCGCGCA-----48
 QY 27 LysGlyValAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlnAlaProAlaProVal 46
 Db 49 -----GGGTTTGGGGGGGATGCTTATCTGAAGCCCGACCGCGAGCTGCT 96
 QY 47 ValGlyValAlaThrValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArg 66
 Db 97 TATATTACGGAACGCTCAGCGCGCGCGCATCAGCGCGCGCTTCTGCACAGGGAG 156
 QY 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyLeuIleGlnLysArg 86
 Db 157 ATTTCCGCGTCCAACTCGTATCGTGGCGCGCGCATCGGCGCGAGATTAGAACTT 216
 QY 87 LeuPheGlnGluSerIleValArgAlaGlyGlnProLeuIleAspSer---105
 Db 217 TATGTCAAACTCGGCGACAGGTTAAAGGCGATTGATTGGGAAATCAATTGCACC 276
 QY 106 -----SerThrTyrluaAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
 Db 277 TCGGAGCAATACCGCTCATACGGAATAATCCAAATTGGAACGATACGCGCAACTG 336
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrlsProLeu 139
 Db 337 GTGTCGGCACAGATTGCTATGGCGACGCGAGAAAGAAATTAAGCTCGCGCGCTTG 396
 QY 140 ValAlaAlaGlnAlaValSerArgGlnLysThrAspAlaValAlaThrAlaLysArgSer 159
 Db 397 TGGAGAGATATGAGCGCTTAAGAAAGATTGGAAGCCACAGATGCTTCGCCGC 456
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 Db 457 GCCAAGACCAATGTTGCGGAGCGTCAATCAGACAGCAAAATTTCCATCAAT 516
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 517 ACCGCCAGTCCGAAATTGGCTACCGCGCATTAACCGCAAGATGCGCGACGCTGGTG 576
 QY 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 577 GCGATTCTCGTGAAGAGGCGACACTGTACACGCGCG-----615
 QY 213 IleArgGlnThrAsnProMetTyrlValAsnValThrGlnSerAlaSerGluValMetLys 232

Db 616 -----CAGTTCAGCGCGACGATGTCATTCAGCGCAAT-----CTGATATCATGTTG 663
 QY 233 LeuArgArgGlnIleLeuAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 664 AACAAATGCAAGATTGCGGAGCGCGATTCACAGGTGAAGCGCGCGCATATTTCG 723
 QY 251 ValGlyIleLysPheAspAspGlyThrValTyrlProGlnLysGlyArgLeuLeuPheAla 270
 Db 724 TTACATATTTCGCAACCGCATACG-----CCGATTAAAGCGAAGCTCGACAGCTC 777
 QY 271 AspProValVal-----AsnGluSerThrGlyGlnIleThr 282
 Db 778 GACCCCGGCTGACCAAGATGCTGCGCGGCTACAAACAGCAGTACGATACGCTTCC 837
 QY 283 -----LeuArgAlaAlaValIleProAsnAspGlnIleLeuMetPro 296
 Db 838 AATCGGCTACTATTATGCGCGTGTGTTGTCGCAATCCGACGCGCAAACTGCCACG 897
 QY 297 GlyLeuTyrlValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
 Db 898 GGGATGACGACGCGAATACGTTGAATCGACGCTGAATAATGCTGATATTTCG 957
 QY 317 GlnGlnAlaVal--ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 958 TCGCTGACCGTGAATAATCGCGCGCGCGCGCGCTTGTGCGCTGTGGGTGCAGACGCG 1017
 QY 336 GlyMetGluProAlaGlnValThrValAlaGlnGlnGlnGlyThrAsnThrIleValThr 355
 Db 1018 AAGCGCGCGAAGCGCAATCCGACCGGTATGAGACAGATGTAATCCGAAGTAA 1077
 QY 356 SerGlyLeuLysAspLysValAlaValGlnGlyIleSerIleAlaGlyIleThr 375
 Db 1078 AGCGGTTGAAGAGGGGGAAGAGTGCATCTCGAATAATACCGCGCGCA--GCAACA 1136
 QY 376 GlyAlaLysLysValThrProLysGluTrrPalSerSerGluAsnGlnAlaAlaPro 395
 Db 1137 GGAAAGCGGGAACGCGCCCTAG-----CGCGCGCGCGCGCGCG 1175

RESULT 25
 AAZ12305 standard; DNA; 1179 BP.
 XX
 AC AAZ12305;
 XX
 DT 08-OCT-1999 (first entry)
 XX
 DE *Neisseria meningitidis* strain A complete ORF85 sequence.
 XX
 KM *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 treatment; *Neisseria* infection; meningitis; septicemia; gonorrhea; ss.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN W0924578-A2.
 XX
 PD 20-MAY-1999.
 XX
 PF 09-OCT-1998; 98MO-IB01665.
 XX
 PR 01-SEP-1998; 98GB-0019016.
 PR 06-NOV-1997; 97GB-0023516.
 PR 14-NOV-1997; 97GB-0024190.
 PR 18-NOV-1997; 97GB-0024386.
 PR 27-NOV-1997; 97GB-0025158.
 PR 10-DEC-1997; 97GB-0026147.
 PR 14-JAN-1998; 98GB-0000759.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Grandi G, Massignani V, Pizsa M, Rappuoli R, Scarlato V;
 DR WPI; 1999-327407/27.

DR P-PSDB; AAY38883.
 XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection
 XX
 PS Claim 9, Page 428, 524pp; English.
 XX
 CC Nucleotide sequences AA211972-212358 represent open reading frames
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode
 CC antigenic proteins (see AAY38849-438944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of *Neisseria* infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.
 XX
 SO Sequence 1179 BP, 306 A, 276 C, 379 G, 218 T, 0 other:
 Alignment Scores:
 Pred. No.: 2,39e-16 Length: 1179
 Score: 308.00 Matches: 114
 Percent Similarity: 43.57% Conservative: 69
 Best Local Similarity: 27.14% Mismatches: 176
 Query Match: 15.26% Indels: 62
 Gaps: 11
 US-09-889-756a-2 (1-412) x AA212305 (1-1179)
 QY 7 LysAlaMetArgAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 7 AAAATGATGAATGATGGCGCTGTTGCGCGCGCGCGCA----- 48
 QY 27 LysGlyLysAspAlaAlaGlnGlyGlnProAlaGlyLysAlaProAlaProVal 46
 Db 49 -----GCGTTGGGGGATGCTTATCTGAAGCCCGAGCCGAGCTGCT 96
 QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 Db 97 TATATTACGGAACGGTCAGGCGCGCGACATCATCGCGAGCTTTCTGCAACAGGGAG 156
 QY 67 LeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeuIleGlnLysArg 86
 Db 157 ATTTCGCGCTCAACCTGTATCGTCGCGCGCGAGCGATCGCAATTAAGAACTT 216
 QY 87 LeuPheGlnGlySerTyrValArgAlaGlyGlnProLeuTyrGlnIleAspSer--- 105
 Db 217 TATGTCAAACTCGGCAACAGGTTAAAGGCGCATTTGATTCGGAAATCAATTCGACC 276
 QY 106 -----SerThrTyrGluAlaAsnLeuGluSerAlaArgAlaGlnLeu 119
 Db 277 TCGCAGACCAATACGCTCAATACGGAATAATCCAAATTGGAACGTAACAGCGAGCTG 336
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeu 139
 Db 337 GTGTGCGACAGATGTGATTGGCGAGCGCGGAAGAAATTAAGCTTCAGCGCGCTTG 396
 QY 140 ValAlaAlaGluAlaValSerArgGlnLysThrAspAlaAlaValThrAlaLysArgSer 159
 Db 397 TGAAGAGTATGACCGCTTAAGAAAGATTGGAAGCGACAGATGCTGCGCCG 456
 QY 160 AlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 Db 457 GCCAAAGCATGTTGCCGAGCTGATCTATACAGACAGCAAAATTTCCATCAT 516
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 517 ACCCCGAGTGAATGGCTACACGCGCATTAACGCAAGATGAGCGACAGTGAGTG 576
 QY 193 GlnSerLysValSerGlyLysThrLeuLeuAsnAlaGlyAspThrValLeuAlaThr 212
 Db 577 GCGATTCTCTGGAAGAGGCGCAACTGTGAACGCGCG----- 615
 QY 213 IleArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLys 232

Db 616 -----CAGTCTACGCGACGATGTCCAAATGGCGAAT-----CTGATATGATGTTG 663
 QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 664 AACAAATGCAATGTGCGAGGAGGATATTAACAAAGTGAAGCGGGCGAGATATTTTCG 723
 QY 251 ValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuPheAla 270
 Db 724 TTTACGATTTTGTTCGCAACCGGATACG-----CCGATTAAAGCGAAGCTGCACAGGTC 777
 QY 271 AspProValVal-----AsnGluSerThrGlyGlnIleThr 282
 Db 778 GACCCCGCGGTACACAGATGCTGCGCGCGCTACAAACAGCATGATACCGCTTCC 837
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 Db 838 AATGCGGTCTACTATTATATGCCCCGTTGTCGCGAATCCGACGCGCAAACTCGCACG 897
 QY 297 GlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValPro 316
 Db 898 GCGATGACGACGCAATACGTTGAATGACAGCGTGTGAATAATGCTGATTAATTCG 957
 QY 317 GlnGlnAlaVal--ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 958 TCGCTGACCGTGAATAATCGCGCGCGCGCGCTTGTGCGCGTGTGGGTGACAGCGCG 1017
 QY 336 GlyMetGluProArgGluValThrValAlaGlnGlnGlnIleThrAsnThrPileValThr 355
 Db 1018 AAGCGCGCGGAACCGCAAAATCCGACCGGTATGAGACAGATGATGATACCGAAGTAAA 1077
 QY 356 SerGlyLeuLysAspGlyAspLysValValIleGluGlyIleSerIleAlaGlyIleThr 375
 Db 1078 AGCGGTTGAAAGGGGGGACAAAGTGTATCTCCAAATTAACCGCCCGCA-GCAGCA 1136
 QY 376 GlyAlaLysLysValThrProLysGluThrPalaSerSerGluAsnGlnAlaAlaPro 395
 Db 1137 GGAAGGCGGAACGCGCCTAG-----CGCCCGCGCGCCCG 1175
 RESULT 26
 AA253631 standard; DNA, 1179 BP.
 ID AA253631
 AC AA253631;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE *Neisseria gonorrhoeae* ORF 290 partial DNA sequence SEQ ID NO:1211.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 OS
 XX *Neisseria gonorrhoeae*.
 PN
 XX MO9957280-A2.
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX

PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 98US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzia M, Rappuoli R, Ratcl G, Scalato E, Scarselli M;
 PI Tectelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR P-PSDB; AAY74871.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 XX Claim 7; Page 678; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of.
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenetic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 1179 BP; 306 A; 276 C; 379 G; 218 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 2,39e-16 Length: 1179
 Score: 308.00 Matches: 114
 Percent Similarity: 43.57% Conservative: 69
 Best Local Similarity: 27.14% Mismatches: 176
 Query Match: 15.26% Indels: 62
 DB: 21 Gaps: 11
 US-09-889-756a-2 (1-412) x AA253633 (1-1179)
 QY 7 LysAlaMetArgAlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGly 26
 Db 7 AAAATGATGAATGGCGCTGTTGCGCGCGCGCGCA----- 48
 QY 27 LysGlyGlyAspAlaAlaGlnGlyGlyGlnProAlaGlyArgGlyAlaProAlaProVal 46
 Db 49 -----GGCGTTGGGGCGGATGCTTATCTGAAGCCGACCCGACGGCTGCT 96
 QY 47 ValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArg 66
 Db 97 TATATTAGCAAGACGGTCAGCGCGCGCGACATCAACCGCGATTCGCAACAGGGAG 156
 QY 67 LeuGlnSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyIleIleGlnLysArg 86
 Db 157 ATTTGCGCGCTCAACCTGTGATCGGTGCGCGCGCATCGGGCGAGATTAAAGAACTT 216
 QY 87 LeuHegInGlnGlySerTyValArgAlaGlyGlnProLeuTyArgInLleAspSer--- 105
 Db 217 TATGTCAAACTCGGGCAACAGGTTAAAGGCGGATTGATTCGGAATCAATTGCAACC 276
 QY 106 -----SerThrTyGlnLalaLeuLeuGlnSerAlaArgAlaGlnLeu 119
 Db 277 TCGCAGACCAATAGCGCTCAATACGAAATAATTCGAATTGGAACGTTATCAGCGAGCTG 336
 QY 120 AlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyLysProLeu 139

Db 337 GTGTGCGACAGATTGATTCGGCAGCGCAGAGAAATATAGCTTCAGCGCGCTTG 396
 QY 140 ValAlaAlaGlnAlaValSerArgGlnGlyTyAspAlaAlaValThrAlaLysArgSer 159
 Db 397 TGGAGGATGATGGACCGCTTAAGAAAGATTGGAAAGCCGACAGATTCGCTTCCGCC 456
 QY 160 AlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsn 179
 Db 457 GCCAAGCAATGTTGCGAGCTGAGAGCTCTATACAGACAGCAAAATTCATCATCAT 516
 QY 180 Arg-----SerArgIleThrAlaProIleSerGlyPheIleGly 192
 Db 517 ACCCGCAGTCGGAATTGGGCTACACCGCATTAACCAAGATGACGACGACGCGTGGT 576
 QY 193 GlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThr 212
 Db 577 GCGATTCTCGTGAAGAGGGGACACTGTGAACCGCGC----- 615
 QY 213 IleArgGlnThrAspProMetTyValAsnValThrGlnSerAlaSerGlyValMetLys 232
 Db 616 -----CAGTCTACGCCGACGATTCGCAATTCGCAAT-----CTGATATGATGTTG 663
 QY 233 LeuArgArgGlnIleAlaGlnGlyLysLeu-----LeuAlaAlaAspGlyValIleAla 250
 Db 664 AACCAAAATGAGATTCGCGAGGGGATTAACCAAGGTGAAGGCGGCGAGATATTGCG 723
 QY 251 ValGlyIleLysPheAspAspGlyThrValTyProGlnLysGlyArgLeuLeuPheAla 270
 Db 724 TTTTCGATTTCCTCCGAACCGGATACG-----CGATTAAAGCGAAGCTGCACAGCTC 777
 QY 271 AspProValVal-----AsnGlnSerThrGlyGlnIleThr 282
 Db 778 GACCCGCGGCTACACAGATGTGTCGGCGGTACAAAGCAGTACGATGATGCGCTTCC 837
 QY 283 -----LeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetPro 296
 Db 838 AATCGGCTACTATATATATGCCCCGTTGCTGCGCAATCCGACGCAAACTTCGCCACG 897
 QY 297 GlyLeuTyValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValPro 316
 Db 898 GGGATGACGACGCGCAATACGATGGAATGACGCGTGTGAATAATGCTGATATTTCG 957
 QY 317 GlnGlnAlaVal--ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGly 335
 Db 958 TCGCTGACCGTGAATAATTCGCGCGCGCGCGCTTGTGCGGCTTGGCGACAGCGC 1017
 QY 336 GlyMetGluProArgGlyValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThr 355
 Db 1018 AAGCGCGCGAAGCGCAAAATCCGACCCGCTATGAGACAGTATGATACCGAATGAAA 1077
 QY 356 SerGlyLeuLysAspGlyAspLysValValValGlnGlyIleSerIleAlaGlyIleThr 375
 Db 1078 AGCGGTTGAAGAGGGGAGCAAGATGTGATCTCCGAATAATACCGCGCGCGA-GCAGCA 1136
 QY 376 GlyAlaLysLysValThrProLysGlyLysThrLysSerSerGlnGlnAlaAlaPro 395
 Db 1137 GGAAGCGCGAAGCGCCCTAG-----CGGCGCGCGCGCGCG 1175
 RESULT 28
 AAS74537/c
 ID AAS74537 standard; cDNA; 3732 BP.
 AC AAS74537;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #10341.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; se.
 XX Homo sapiens.
 OS

PS Claim 1; SEQ ID No 28696; 103bp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3732 BP; 833 A; 1010 C; 961 G; 928 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.88e-15 Length: 3732
 Score: 304.50 Matches: 73
 Percent Similarity: 51.22% Conservative: 53
 Best Local Similarity: 29.67% Mismatches: 83
 Query Match: 15.08% Indels: 37
 DB: 23 Gaps: 5
 US-09-889-756a-2 (1-412) x AAS92892 (1-3732)
 QY 184 ThrAlaProIleSerGlyPheIleGlyIleSerIleValSerGlyIleThreuleuAsn 203
 Db 2562 ACCTTCCCGGTCAACGGGGAATTCACCCGCAATGGAATCTTCAATGCTGGAC 2503
 QY 204 AlaGlyAspThrValIleuAlaThrIleArgIleThrAsnProMetIleVal----- 221
 Db 2502 -----CTTTCGAAACCACTGGCTCATATGCTCCGAAA 2470
 QY 222 -----AsnValThrGlnSerGluValMetIleuValArg 235
 Db 2469 TCCATATGGAATCTCTGCTCTCTACAGATCTCCACGACTTCTGCGCTGAAACAG 2410
 QY 236 GlnIleAlaGlyIleValIleuAlaIleAspGlyValIleAlaValGlyIleAspHe 255
 Db 2409 GAACCTGGGAATGGACGCTGAACAGAAAGCAAGCAAGTCTCACTGATCAC 2350
 QY 256 AspAspGlyThrValIleProGlyIleValGlyIleuPheIleAspProValValAsn 275
 Db 2349 AGTACCGCATTAATGTTCCCGACGACGTTACGCTGAAATCTCTGACGTTACCGTTGAT 2290
 QY 276 GluSerThrGlyGlnIleThrIleuAlaIleValIleProAsnAspGlnAsnIleuMet 295
 Db 2289 CAGACCACTGGCTCTATACCTTACCGCGCTACTTCCCGAACCCGATCACACTCTGCTG 2230
 QY 296 ProGlyIleuValIleValIleuMetAspGlnValIleAlaIleAspAsnAlaPheValVal 315
 Db 2229 CCGGGTATGTTCTGTCGCGGACGCTGGAAGAAGGGCTTATCCAAACGCTATTATTGTC 2170
 QY 316 ProGlnGlnAlaValIleThrArgIleAlaIleAsp-----ThrValMetIleValAsnAla 333
 Db 2169 CCGGAACAGGGCGTAAACCGTACCGCGGTGCGCATGACCGCACTGCTAGTGGCGCG 2110
 QY 334 GlnGlyIleMetGluProArgGluValThrValIleGlnGlnGlnIleThrAsnTrpIle 353
 Db 2109 GATGACAAAGTGAACCCGCTCCGATCTTCCAAAGCCGCTATTGGGATTAAGTGGCTG 2050
 QY 354 ValThrSerGlyIleuValAspGlyAspIleValValIleGlyIle----- 369

Db 2049 GTGACAGAAAGTCTGAAAGCAGCGCATCGCTAGTATAGTGGCGTCAAGAAAGTCCGT 1990
 QY 370 -----SerIleAlaGlyIleThrGlyIleValIleVal 360
 Db 1989 CCGGTCTCCAGGCGCCCGGTTCCGCAATACCCGACTGAGCGCA----- 1942
 QY 381 ThrProGlyIleuThrAlaSerSerGluAsnGlnAlaIleAlaIleProGlnSerGlyValGln 400
 Db 1941 ---CCGAGCGAAATCGCGCGTGCAGAAAAGCACTGAGCGCGCCCGCACAGCAACAGGG 1885
 QY 401 ThrAlaSerGluAlaIleVal 406
 Db 1884 GGGGACACACAGCAAGCCAG 1867
 RESULT 30
 AAS93885/C
 ID AAS93885 standard; cDNA; 3732 BP.
 AC AAS93885;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29689.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001NO-US08631.
 PR 31-MAR-2000; 2000US-0540217;
 PR 23-AUG-2000; 2000US-0649167;
 XX
 PA (HSE-) HYSER INC.
 XX Drmanac RT, Liu C, Tang YT;
 P1
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG29698.
 DR
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 29689; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

SX Sequence 3732 BP; 833 A; 1010 C; 961 G; 928 T; 0 other;

Alignment Scores:

Pred. No.:	1.88e-15	Length:	3732
Score:	304.50	Matches:	73
Percent Similarity:	51.22%	Conservative:	53
Best Local Similarity:	29.67%	Mismatches:	83
Query Match:	15.08%	Indels:	37
DB:	23	Gaps:	5

US-09-889-756A-2 (1-412) x AAS93885 (1-3732)

```

QY 184 ThrAlaProIleSerGlyPheIleGlyIleSerIleValSerGlyIleThrLeuLeuAsn 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2562 ACCCTCCCGGTCAACGGCGAATTCACCCCAATGTAATCTTCAAAATGTCGGAC 2503
QY 204 AlAGlyAspThrValLeuAlaThrIleArgGlnThrAsnProMetIleVal----- 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2502 -----CTTTCGAAACACCTGGCCCTATATGTCGGAAA 2470
QY 222 -----AsnValThrGlnSerAlaSerGluValMetIleValArg 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2469 TCATATGGAATTCCTGGTCACTTACAGGATCTCCAAACGCTTCTCGCGCTGAAACAG 2410
QY 236 GlnIleAlaGluGlyIleLeuAlaAlaAspGlyValIleAlaValGlyIleValPhe 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 GAATCGCCGAATGCGACGCTGAAACAGAGAACGCGAAACCAAGTGTACATCACC 2350
QY 256 AspAspGlyThrValIleValProGluValGlyValLeuLeuValAspProValValAsn 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2349 AGTACGCGCATTAAGTTCCTCCGACGACGCTGACCTGGAATTCCTGACGTTACGTTGAT 2290
QY 276 GluSerThrGlyIleIleThrIleValAlaValProAsnAspGlnAsnIleLeuMet 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2289 CACACCACTGGGCTATACCTACCGCTATCTTCCCGAACCGGATCACACTGCTGCTG 2230
QY 296 ProGlyLeuThrValArgValLeuMetAspGlyValAlaValAspAsnAlaPheValVal 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2229 CCGGGATATGTCCTGCGCGACGCTGGAAGAAAGGCTTATCCAAACGCTATTTAGTC 2170
QY 316 ProGlnGlnAlaValThrArgGlyAlaValAsp-----ThrValMetIleValAsnAla 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2169 CCGCAACAGGGCGTAAACCGCTAGCCGCGGGGATGCCACCGTACGTGATTTGGGCGC 2110
QY 334 GlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTrpIle 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2109 GATGACAAAGTGAACCCGCTCGATGCTTGCACAGCAGCGCTATTGGCGATAGTGGCTG 2050
QY 354 ValThrSerGlyLeuIleValAspGlyValAspValValValGlnGlyIle----- 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2049 GTATACGAAGGTCTGAAGACAGCGGATCGGTATATAGTGGCTGCAAAAGTGGCT 1990
QY 370 -----SerIleAlaGlyIleThrGlyAlaValValVal 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1989 CCGTGTGTCCAGGCGCCCGGCTTCCGCGACATACCGACCTAGCGGA----- 1942
QY 381 ThrProIleGlnTrpAlaSerSerGluAsnGlnAlaAlaProGlnSerGlyValGln 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1941 ---CCAGCGCAAAATCCGCGCGTGCAGAAACCACTAGCGCGCCGACCAAGAGGCG 1885
QY 401 ThrAlaSerGluAlaVal 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1884 GGGGACACACACCAAGCAG 1867

```

RESULT 31

AAA81495

ID AAA81495 standard; DNA; 44608 BP.

AC AAA81495;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_42 SEQ ID NO:42.

DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Maignant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappelli R, Pizzi M;

XX MPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisseria infections, for example, N.gonorrhoea -
 PS Claim 7; Page 1283-1296; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AA825620 to AA825663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisseria bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX Sequence 44608 BP; 10938 A; 10835 C; 11999 G; 10834 T; 2 other;

XX Alignment Scores:

Pred. No.:	3.79e-14	Length:	44608
Score:	304.50	Matches:	117
Percent Similarity:	44.34%	Conservative:	67
Best Local Similarity:	28.19%	Mismatches:	176
Query Match:	15.08%	Indels:	56
DB:	21	Gaps:	13

US-09-889-756A-2 (1-412) x AAA81495 (1-44608)

```

QY 25 CysGlyIleValGlyIleValAlaGlnGlyIleValProAlaGlyArg----- 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 TCGCGC-----GATCGCGCGCGACGCGTGTGGGCGGATGCTTATCT 574

```



```

Qy 41 GluAlaProAlaProValValGlyValVal---ThrValHisProGlnThrValAlaLeu 59
Db 575 GAACCCGAGCGCGAGCTGCTTATATACGAAACGGTCAGCGCGGACATCAGCCGG 634
Qy 60 ThrValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnVal 79
Db 635 ACGGTTCTCGACACAGGAGATTTCCGCGTCAACCTGATTCGTCGCGCCAGCA 694
Qy 80 GlyGlyIleIleGlnLysArgLeuPheGlnGluGlySerTyrValArgAlaGlyGlnPro 99
Db 695 TCGGGGAGATTAAGATACCTTATATGTCAACTCGGGCAACGATTAAGAGGCGATTG 754
Qy 100 LeuTyrGlnIleAspSer-----SerThrTyrGluAlaAsnLeu 112
Db 755 ATTGGGAAATCAATTGCACTCGACGCAACCAATCGCTCAATACGAAATTCACAGTTG 814
Qy 113 GluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAsp 132
Db 815 GAAACGTATCAGCGGAGAGCTGTCGCGACAGATTGCGACGCGCGAGACGAA 874
Qy 133 LeuAlaArgTyrLysProLeuValAlaAlaGlnAlaLysArgGlnGluTyrAspAla 152
Db 875 TATAAGCGTCAAGCGCGCTTATGAGAGAAAACGCGACTTCAAGAGAGATTGGAAGC 934
Qy 153 AlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaIleLys 172
Db 935 GCGCAGAGATGCGTTTCCCGCCGCAAAACCAATGTCGCGAGCGAGGCTTATATCGA 994
Qy 173 SerAlaGlyIleAsnLeuAsnArg-----SerArgIleThrAla 185
Db 995 CAGAGCAAAATTTCCATCAATATACCGCCGATTCGCGATTCACGCGATTATCCGCA 1054
Qy 186 ProIleSerGlyPheIleGlyGlnSerLysValSerGlyGlnLeuLeuAlaGly 205
Db 1055 ACGATGAGCGGCGAGTGTGCGCATTTCTGTGAAGAGGGGAGAGCTGTGAACGCGCG 1114
Qy 206 AspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyrValAsnValThrGln 225
Db 1115 -----CAGTCACCGCGAGATTCCTCAATTCGCCAAT 1147
Qy 226 SerAlaSerGluValMetLysLeuArgArgGlnIleAlaGlnGlyLysLeu-----Leu 243
Db 1148 -----CTGATATGATGTTGACAAACAAATGCAATGCGCGGCGATATATCAAGGTG 1201
Qy 244 AlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgIleThrValTyrProGlu 263
Db 1202 AAGCGGGGCGAGATATTTCTTTACGATTTTGTCCGAACCGGATACG-----CCGATT 1255
Qy 264 LysGlyArgLeuLeuPheAlaAspProValVal-----Asn 275
Db 1256 AAGCGAGAGCTCGACAGCTGACACCCCGGCTGACACAGATGTCGTGGCGGTTTACAC 1315
Qy 276 GluSerThrGlyGlnIleThr-----LeuArgAlaAlaValProAsn 289
Db 1316 AGCAGTACGAGTACGCGCTTCAATGCGGTCTACTATTATTCGCCGCTGTTGCGCAT 1375
Qy 290 AspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaVal 309
Db 1376 CCGAGCGGCAAACTCGCACCGGAGTACGACGAGATACGATTGAATCGACGCGCTG 1435
Qy 310 AspAsnAlaPheValValProGlnAlaVal---ThrArgValAlaLysAspThrVal 328
Db 1436 AAAAATGCTGATTTATTCGTCGTCGACCGTGAAGAAATCCGGGGCGCAAGCGCTTTG 1495
Qy 329 MetIleValAsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGln 348
Db 1496 CGCGTGTGGGTGGGAGCGAGCGGCGGAGCGCAATCCGAGACCGGATAGAGAC 1555
Qy 349 GlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspValValValGlnGly 368
Db 1556 AGTATGATACCGAATAAAAGCGGTTGAAGAGGCGGCAAAAGTGTATCTCCGAA 1615

```

```

Qy 369 ILeSerIleAlaGlyIleThrGlyAlaLysValThrProLysGluTrpAlaSerSer 388
Db 1616 ATAAACCGCCCGCA-GCAACAGAAAGCGGCAACGCGCTTACGCGCGCGCGCG 1674
Qy 389 GluAsnGlnAlaAlaAla---ProGlnSerGlyValGlnThrAla 402
Db 1675 ATAAACGAATATGCGCTGTAACACGGAACGGTTTTCAGACGCA 1719

RESULT 32
AAF28554
ID AAF28554 standard; DNA; 269223 BP.
XX
AC AAF28554;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #41.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
EN W0200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000MO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
WP1; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
XX
PS Claim 1; Page 486-545; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 269223 BP; 77067 A; 56596 C; 57380 G; 78180 T; 0 other;

Alignment Scores:
Pred. No.: 3.34e-13 Length: 269223
Score: 304.50 Matches: 85
Percent Similarity: 54.42% Conservative: 32
Best local Similarity: 39.53% Mismatches: 73
Query Match: 15.08% Indels: 25
DB: 22 Gaps: 6

US-09-889-756a-2 (1-412) x AAF28554 (1-269223)
Qy 21 ValLeuSerSerGlyLysGly-----GlyAspAlaAlaGlnGlyGln 36
Db 238995 CTGCTCAACCCCTTTGGCAACGAGATATACCTTGGGTGGGTACTACCGAGGTGTA 239054
Qy 37 ProAlaGlyArg-----GluAlaPro----- 43

```


Db 23905 CTCACGGGCGATGTATGGATGATCCAAATCAAGACACCGCTCCAAAGAGCTCCCA 239114
Qy 44 -----AlapProValValGlyValValThyValHisProGlnThrValAlaLeuThr 60
Db 239115 AACCAAAACGACGACGATGATGCGGTGGAAGCGGTATCCCAATCATGATCAATCAATCA 239174
Qy 61 ValGluLeuPro-----GlyArgLeuGlnSerLeuArgThyAlaAspValArgAlaGln 78
Db 239175 AATGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 239234
Qy 79 ValGlyValGlyLeu-----GlnGlnArgLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGln 97
Db 239235 ATCAGCGGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 239294
Qy 98 GlnProLeuValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 117
Db 239295 CAAAGTGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239354
Qy 118 GlnLeuValThyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 137
Db 239355 GAGCTGTCACAGCGGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239414
Qy 138 ProLeuValAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 157
Db 239415 CCATTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239474
Qy 158 ArgSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 177
Db 239475 CCGGCAAGTGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239534
Qy 178 LeuAsnArgSerArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 197
Db 239535 CAGTAGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239594
Qy 198 GlnGlyThyLeuLeuAsnAlaGlyAspThyThyValLeuAlaThr 212
Db 239595 GTTGTGTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239633
RESULT 33
ABO24964/c
ID ABO24964 standard; DNA, 607 BP.
AC ABO24964;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 11555.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; de.
XX Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIC-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Gueutig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.
PS This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridized to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 607 BP; 121 A; 53 C; 159 G; 274 T; 0 other;
XX
Alignment Scores:
Pred. No.: 2,24e-15 Length: 607
Score: 292.00 Matches: 60
Percent Similarity: 56.00% Conservative: 52
Best Local Similarity: 30.00% Mismatches: 78
Query Match: 14.46% Indels: 10
DB: 24 Gaps: 2
US-09-889-756a-2 (1-412) x ABO24964 (1-607)
Qy 185 AlapProLeuSerGlyPheLeuGlnSerGlyValSerGlnGlyThyLeuAsnAla 204
Db 607 TCTCCGATTAAGATGCGCATTAATTAATGCAAGTAACGAAACGATTAAATCAAAAC 548
Qy 205 GlyAspThrThyValLeuAlaThrLeuGlnThrAspProMetCysValAsnValThr 224
Db 547 GATCAACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 488
Qy 225 GlnSerAlaSerGlnValMetCysLeuArgGlnGlnLeuGlnGlyLeuLeuAla 244
Db 487 CAATCCAAACGACCTCTCAAGCTTAACCAAACTTAACGAATTAACGCTTAACCAAA 428
Qy 245 AlaAspGlyValLeuAlaGlyLeuPheAspAspGlyThyValTyProGlnGly 264
Db 427 AAAAAGCAAAACCAAAATATGATCAATCAACCAATTAACGATTAATTCGCAAAAC 368
Qy 265 GlyArgLeuLeuPheAlaAspProValValAsnGlnSerThyGlnGlnLeuArg 284
Db 367 GATACCTTAATTAATCTTAACGTTACGTTAAACCAACCTTAATCTTAACCTTAACG 308
Qy 285 AlaAlaValProAsnAspGlnAsnLeuMetProGlyLeuValValArgValLeuMet 304
Db 307 GCTATCTCCCGAACCCGATCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 248
Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThyArgGlyAla 324
Db 247 AAAAAGCAAAACCAAAATATGATCAATCAACCAATTAACGATTAATTCGCAAAAC 188
Qy 325 LysAsp-----ThyValMetLeuValAsnAlaGlnGlyCysLeuProArgGlyVal 342
Db 187 CGTAACGATCAACCGCTCAATTAATTAACGCAATTAACCAATTAACCAACCGCTCAATC 128
Qy 343 ThrValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 362
Db 127 GTTCAACCAACCAATTAATTAACGATTAATTAATTAATTAATTAATTAATTAATTAAT 68
Qy 363 LysValValValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 382

Db 67 CGCGTAATAATAATACTA-----CAAAATAATCGTCT 32

RESULT 34

ID ABQ24965 standard; DNA; 607 BP.

AC ABQ24965;

DT 12-JUN-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 11556.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

OS Homo sapiens.

PN MO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIDENOMICS AG.

PI Olek A. Piepenbrock C, Berlin K, Guelzig D;

DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.

CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the

CC method for determining the degree of cytosine methylation described in

CC the disclosure of the invention.

XX Sequence 607 BP; 274 A; 159 C; 53 G; 121 T; 0 other;

XX

Alignment Scores:

Pred. No.: 2.24e-15 Length: 607

Score: 292.00 Matches: 60

Percent Similarity: 56.00% Conservative: 52

Best Local Similarity: 30.00% Mismatches: 78

Query Match: 14.46% Indels: 10

DB: 24 Gaps: 2

US-09-889-756a-2 (1-412) x ABQ24965 (1-607)

Qy 185 AlaProIleSerGlyPheIleGlyGlnSerIleValSerGlyGlyThrLeuLeuAsnAla 204

Db 1 TCTCCGATTACAGATGCCATTAAATGACAGCTAACGAAACCGATTAAATACAAAC 60

Qy 205 GlyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetCylTrValAsnValThr 224

Db 61 GATCAAGAGACTACGCTACCAACCGTACCAACCACTTAATCGATTAAATATAC 120

Qy 225 GlnSerAlaSerGlyValMetIleValArgArgGlnIleAlaGlyIleValLeuAla 244

Db 121 CAATCCACACAGCACTCTTCTACCGCTTAAACAAACAACTACGAAATACACGCTTAAACAA 180

Qy 245 AlaAspGlyValIleAlaValGlyIleValPheAspAspGlyThrValTrpProGlyLys 264

Db 181 AAAAAGACAAACCAAAATATCATCATATACCAATATACCAATATATATATATATATATATAT 240

Qy 265 GlyArgLeuLeuPheAlaAspProValValAsnGlySerThrGlyGlnIleThrLeuArg 284

Db 241 GATACGCTAAATCTCTTAACGTTACCGTTAAATCAACCACTTAATCTATACCGCTTAC 300

Qy 285 AlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuThrValAlaValLeuMet 304

Db 301 GCTATCTTCCGAAACCGAATCACTCTACTACCGAATATATTCGATACCGCACTCTA 360

Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAla 324

Db 361 AAAAATACTTATCCAAACCGTATTTATATCCGCAACAAACGTTACCGCACTCTA 420

Qy 325 LysAsp-----ThrValMetIleValAsnAlaGlnGlyIleMetGlyProArgGlyVal 342

Db 421 CGTAAACGATCCACCGTAAATTAATTAACGCAATTAACCAATTAACCAATTAACCAATTAAC 480

Qy 343 ThrValAlaGlnGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuLysAspGlyAsp 362

Db 481 GTTACAAACCAAACTATTAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540

Qy 363 LysValValValGlyGlyIleSerIleAlaGlyIleThrGlyAlaLysLysValThrPro 382

Db 541 CGCGTAATAATAATACTA-----CAAAATAATCGTCT 576

RESULT 35

ID AAT42063 standard; DNA; 1830121 BP.

AC AAT42063;

DT 14-SEP-1999 (first entry)

XX Haemophilus influenzae complete genome sequence.

KW Genome; bacterium; Haemophilus influenzae; computer readable medium;

KW expression modulating fragment; regulation; gene expression; vector;

OS Haemophilus influenzae.

PN WO9633276-A1.

PD 24-OCT-1996.

PP 22-APR-1996; 96WO-US05320.

PR 07-JUN-1995; 95US-0487429.

PR 21-APR-1995; 95US-0426787.

PR 07-JUN-1995; 95US-0476102.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (UYJO) UNIV JOHNS HOPKINS.

PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

XX WPI; 1996-485782/48.

XX Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching

Claim 1, Page 77.2-77.1091, 1291pp; English.

CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
CC sequence at least 99% identical to (1). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.

XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:

Pred. No.:	1,14e-10	Length:	1830121
Score:	286.00	Matches:	105
Percent Similarity:	41.01%	Conservative:	66
Best Local Similarity:	25.18%	Mismatches:	148
Query Match:	14.17%	Indels:	98
DB:	17	Gaps:	14

US-09-889-756a-2 (1-412) x AAT42063 (1-1830121)

QY 6 PhelysalawetargalialaialaaleuallalavalaleuallleuserSerCys 25
DB 946822 TTTAATATGATTAAGCGTAATAGCCGCAATTCGA----- 946863
QY 26 GlylyseGlyleypalaialaglnGlyGlnProCaglylaAglylaProAlaPro 45
DB 946864 -----GCAATGCCA-----GATCTTCACAGCCA 946887
QY 46 ValValGlyValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGly 65
DB 946888 GTAAAC--GCACTTGAAGTTCAACGCGTAATAGCAAGCCGCAATTAACACAAACAGGT 946944
QY 66 ArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGlyGlyLeileGlnLys 85
DB 946945 CTTGTGGTCCAAATCAAGGCGCATGCTCAGTACACAAATGCGGCGCGTTTCACAA 947004
QY 86 ArgLeuPheGlnGlySerTyrrValArgAlaGlyGlnProLeuTyrrGlnLleAspSer 105
DB 947005 GTACTTGTCAAAATGACAAATATGTAAAAAGGTGAGGTGTTGTGAGCTTGATAGT 947064
QY 106 SerThrTyrrGlnAlaGlnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThr 125
DB 947065 TCTGTTCACAGCATTAATTCACAGCTGCTCAGACCAATATACGACTTCGCAACT 947124
QY 126 LeuAlaLysAlaAspAlaAspLeuAlaArgTyrrSerProLeuValAlaAlaGlnAlaVal 145
DB 947125 -----TACCAACGTTATGCGTTTATTAATACCAATGCTGTA 947163
QY 146 SerArgGlnGlnTyrrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLys 165
DB 947164 TCACGTCMAAGAAATGATACGCAAAAGCGCTTATGATGCTCAAGTACGATGATGAA 947223
QY 166 AlaAlaGlnAlaAlaLysSerAlaGlyLleAsnLeuLysnArgSerArgLleThrAla 185
DB 947224 TCTCTAAGACGCAATTCGA-----CGTCGTAAATTTGTTGCG 947262
QY 186 ProLleSerGlyPheLleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGly 205
DB 947263 CCAATTGATGCAAGACAGATATTTGTAATAATCAATTTGACATATGTCATGTTGGA 947322
QY 206 AspThrThrValLeuAlaThrLleArgGlnThrAsnProMetTyrrValaLeuValThnGln 225

DB 947323 -----ACAGAAATTTGCGCTGTAGAAAGATACAGCTCAATGAAAGCGATTTGCTCTT 947376
QY 226 SerAlaSerGlnValMetLysLeuArgGlnLleAlaGlnGlyLysLeuLeuAlaAla 245
DB 947377 TCACAAATATGATTTAGTAAATTCAT-----ATCGTTCAGCGGCTTACAGCG 947424
QY 246 AspGlyValLleAlaValGlyLysPheAspAspGlyThrValTyrrProGluLysGly 265
DB 947425 ACAACAGATCGCTCGTGGCGCAAAACATT-----TCAGCT 947460
QY 266 ArgLeuLeuPheAlaAspProValAlaAsnGluSerThrGlyGlnLleThrLeuArgAla 285
DB 947461 CGAATCACTCGCATTTGAACTGCCATTAAATTCATCAACAGGTTTGAATGTTACAGCT 947520
QY 286 Alaval---ProAsnAspGlnAsnLleLeuMetProGlyLeuTyrrValArgValLeuMet 304
DB 947521 ACTTTTATCTTGAAGATGGGCATTAATTCCTTCAGGATATGTTCTTCGTTACGCAAT 947580
QY 305 AspGlnValAlaValAspAsnAlaPheValAlaProGlnGlnAlaValThr----- 321
DB 947581 GCACTTCCACTGAACAAATCAAGTTGCTTCCACAACTAGCTATTAAGTCAATATG 947640
QY 321 ----- 321
DB 947641 TATGCGCAAAATTCCTATTACTTGAACCATTAATTCGAAAGCAAGAAAAAATGTCATCA 947700
QY 322 -----ArgGlyAlaLysAspThrValMetLleValAsn 332
DB 947701 GGTAATGAAAAAATTTGCATCGCTCTATCGCGAAACGATCACCGATTTTACTTAAGAT 947760
QY 333 AlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlyThrasnTrp 352
DB 947761 CGTCAAGT-----GTTATGCTCAATTAACGGGAATGAA--- 947796
QY 353 ileValThrSerGlyLeuLysAspGlyAspLysValValValGlnGlyLleSerLleAla 372
DB 947797 -----GTTAAAGTGGAGATTAATTAATTAACAGCGGTACAGCA--- 947835
QY 373 GlyLleThrGlyAlaLysLysValThrProLysGluThrAlaSerSerGlu 389
DB 947836 GGTATTTGTAATGAAAGCTTTGTG-----GAATGATTAATAAAAGAC 947877
DB 947877 -----GAATGATTAATAAAAGAC 947877
RESULT 36
ID ABO90313 standard; DNA; 1086 BP.
AC ABO90313;
AC XX
DT 01-OCT-2002 (first entry)
DT XX
DE M. capsulatus gene #298 for DNA array.
XX.
KW Micro array; gene; ds; differential expression; gene expression.
KW OS
OS Methylococcus capsulatus.
PN W0200255655-A2.
PN 18-JUL-2002.
PD 14-JAN-2002; 2002WC-NO00019.
PF 12-JAN-2001; 2001NC-0000235.
PR 12-JAN-2001; 2001NO-0000239.
PR XX
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX (TIGR-) TIGR.
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillhaug JR, Lossius I, Eisen JA, Frazer CM, Durkin AS;
PI Salzborg SL;
XX WPI; 2002-557818/59.

```

XX Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprising polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes
XX
PS Claim 19; Page 193; 678bp; English.
XX
CC The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC AB090016-AB091855 represent M. capsulatus genes for use in arrays of the
CC invention.
XX
SQ Sequence 1086 BP; 363 A; 278 C; 232 G; 213 T; 0 other;
Alignment Scores:
Pred. No.: 4.88e-14 Length: 1086
Score: 279.50 Matches: 89
Percent Similarity: 45.08% Conservative: 76
Best Local Similarity: 24.32% Mismatches: 162
Query Match: 13.84% Indels: 39
DB: 24 Gaps: 9
US-09-889-756a-2 (1-412) x AB090313 (1-1086)
OY 18 ValAlaLeuValLeuSerSerCysGlyLysGlyGlyAlaAlaGlnGlyGlyGlnPro 37
Db 34 ATTGCCATGTCGTTGCTTGCCTGCAGAGACTGACGACGAAACCGCCAGAAACTTCCA 93
OY 38 AlaGlyArgGlyAlaProAlaProValValGlyValValThrValHisProGlnThrVal 57
Db 94 AAA---TACGAGGACACCCACCCATTTGCAAGACATATCAATATAAAGAAATATGTA 150
OY 58 AlaLeuThrValGluLeuProGlyArgLeuGlnSerLeuArgThrAlaAspValArgAla 77
Db 151 TGT-----CAGATCACGCGCATCAGACATCATCATGAGAGTGAAGGCC 189
OY 78 GlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlyGlySerTyrValArgAlaGly 97
Db 190 ATGGAAGAAAGGCTACTGTCAGACATATATGTCATGAGGGGAGTTGTTATCATCAGGGC 249
OY 98 GlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaLeuGlnSerAlaArgAla 117
Db 250 CAGCGCATGTTCAAAATATATCCAGCCATCTAT-----CTACCA 288
OY 118 GlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLys 137
Db 289 GAATCATGTAAGGGCGAAGCCGAAAGCCAAATCATGATATGATATACAAAACCCAAA 348
OY 138 ProLeuValAlaAlaGluAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLys 157
Db 349 GGCTTGGCGCAGCAAAATATGTTATCGCAGAACGAACTC----- 387
OY 158 ArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsn 177
Db 388 GCTCTGGCCAAAGCCTGACAGCAGCAATGCTGCAAGTAACTGGCAGACCTAT 447
OY 178 LeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSer 197
Db 448 TTGAGTTTCACCGCATTTAAGCCCATTCATGATGATGATGATCATCTCTGCGCCGA 507
OY 198 GluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAsn 217
Db 508 AATGGCAGCCTTTTGCAGGAAGA-----ACACTATTAACACGCTCTCCGATATCAGC 561
OY 218 ProMetTyrValLeuValThrGlnSerAlaSerGluValMetLysLeuArgArgGlnIle 237
Db 562 AAGTTATGGGCTTATTTCAACGTAACCGCAATCTGATATCTCGACTACAAACCAATAC 621
OY 238 AlaGluGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAsp 257

```

```

Db 622 AAAGACACGACGACATCTCTG-----AAAGTCAAGCTGAAATAATGCGCAAT 666
OY 258 GlyThrValTyrProGluLysGlyArgLeuLeuPheAlaAspProValValAsnGluSer 277
Db 667 GGCGCGATTTTCATCTCATCGAAGGAGATGCAAAACGATAGACAGATTTCCGACCAACC 726
OY 278 ThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGly 297
Db 727 GCTGGCAACATCGAGTTTCAGGGGCACTTTCCGAAATCCCAACGCTCTCTCAGACATGGA 786
OY 298 LeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGln 317
Db 787 GAAACCGGAAACCATCTAAAGAAAG---CCATACAAAACCTTGCATCCATCCCA 843
OY 318 GlnAlaValThrArgGlyAla---LysAspThrValMetIleValAsnAlaGlnGly 336
Db 844 AAAGCCACCTTCCAGATGCTCGACAGATGTTGTCTACGTCATCAACATGAAGGCTC 903
OY 337 MetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsnTyrIleValThrSer 356
Db 904 GTCGACAGAGGAGCATCATGATTTGAACGCAAAATCCCATATTTCTTCTGTAAAGCAA 963
OY 357 GlyLeuLysAspGlyAspLysValValGluGlyIleSerIleAlaGlyIleThrGly 376
Db 964 GCGCTCAAGATGACGACAAAGATCTTGATCGAGGGCTTG----- 1002
OY 377 AlaLysLysValThrPro 382
Db 1003 ---CGTAAAGTACATCCT 1017
RESULT 37
AAZ12304
ID AAZ12304 standard; DNA; 1005 BP.
AC AAZ12304;
XX
XX 08-OCT-1999 (first entry)
DE Neisseria meningitidis partial ORF65 sequence.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
XX
XX Neisseria meningitidis.
OS
XX
XX W09924578-A2.
FN
XX
XX 20-MAY-1999.
PD
XX
XX 09-OCT-1998; 98MO-IB01665.
PE
XX
XX 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
PA
XX
XX (CHIR-) CHIRON SPA.
PI
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
XX P-PSDB; AAY38882.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 9; Page 427; 524bp; English.
XX
XX Nucleotide sequences AAZ11972-212358 represent open reading frames

```

CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode
 CC antigenic proteins (see AAY3499-Y38944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of *Neisseria* infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.

XX Sequence 1005 BP; 271 A; 233 C; 315 G; 186 T; 0 other;

Alignment Scores:

Pred. No.:	2.23e-13	Length:	1005
Score:	271.00	Matches:	100
Percent Similarity:	45.45%	Conservative:	60
Best Local Similarity:	28.41%	Mismatches:	141
Query Match:	13.42%	Indels:	52
DB:	20	Gaps:	10

US-09-889-756a-2 (1-412) x AA212304 (1-1005)

```

Qy 75 ValArgAlaGlnValAlaGlyIleIleGlnIleArgLeuPheGlnGlnGlySerThrVal 94
Db 7 GTGGCGCCGCGAGCATGCGGCGAGATTAACTATTATGTCACAACTCGGCAACAGT 66
Qy 95 ArgAlaGlnProLeuTyrglnIleAspSer-----SerThr 107
Db 67 AAAAAGGCGATTGTCGGAATCAATTCGACCTCGACAGCAATACGCTCAATACG 126
Qy 108 TyrGlnAlaSerLeuGlnUserAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAla 127
Db 127 GAAAATCCAAAGTGGAAAGCTATCAGCGAAGCTGTGTGCGACAGTTCATTGGGC 186
Qy 128 IysAlaAspAlaAspLeuAlaArgTyrglnProLeuAlaAlaGlnAlaValaSerArg 147
Db 187 AGCGCGGAGAGAATTAAGCGTCAGCGCGCTTATGAGAGAAACGCACTCCAA 246
Qy 148 GlnGlnTyrAspAlaAlaValThrAlaIleArgSerAlaGlnAlaGlyValIleAlaAla 167
Db 247 GAGGATTGGAAAGCGCGAGATGCTTGGCCGCCCAAGCCAAATGTTGCCGAGCTG 306
Qy 168 GlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg----- 180
Db 307 AAGGCTTAATCAAGACAGCAAAATTTCCATCAATCCCGCAGTCGAAATGGGCTAC 366
Qy 181 SerArgIleThrAlaProIleSerGlyPheIleGlnIleSerIleValSerGlnGlyThr 200
Db 367 ACGGCAATTAACCGGACGATGAGAGCGAGCGGTGCGCATTCCTCGGAGAGGGGCGAG 426
Qy 201 LeuLeuAsnAlaGlnIleAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTyr 220
Db 427 ACTGTGAACCGGCG-----CAGTCTACCGGAGCAT 459
Qy 221 ValAsnValThrGlnSerAlaSerGlnValMetIleLeuArgArgGlnIleAlaGlnGly 240
Db 460 GTCCAAATGGCGCAT-----CTGGATATGATGTGAACAAATGCGAGATGGCGAGGC 513
Qy 241 LysLeu-----LeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspArgGly 258
Db 514 GATATTACCAAGGTGAAGCGGCGGCGAGATATTTCTGTTTCGATTTTGTCCGAACCGGAT 573
Qy 259 ThrValTyrProGlnTyrGlnIleLeuLeuPheAlaAspProValVal----- 274
Db 574 ACG-----CGATTAAAGGAGCTCGACAGCGTGCACCGCGGCGGACACAGATGCG 627
Qy 275 -----AsnGlnSerThrGlnIleThr-----LeuArg 284
Db 628 TCGGCGCGATTACACAGCATGACGATACGCTTCCAAATCGGCTTACTATTATGCGCGT 687
Qy 285 AlaAlaValProAsnAspGlnAsnIleLeuMetProGlnIleuTyrglnValArgValLeuMet 304
Db 688 TCGTTGTGCGGAATCCGAGCGGCAACTCGCCACGCGGAGATGACGACGAATACGGTT 747
Qy 305 AspGlnValAlaValaAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgGly 323

```

```

Db 748 GAAATCGACGCGCTGAAAAATGTGCTGATTATTCGTCGCTGACCGTGAAAAATCGCGGC 807
Qy 324 AlaIysAspThrValMetIleValaAsnAlaGlnIleGlyMetGlnProArgGlnValThr 343
Db 808 GGCAGAGCGCTTTGTGCGCGGTGGTGGCGGACGCGGAGCGCGGCAACCGCAATCCG 867
Qy 344 ValAlaGlnGlnGlnIleThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLys 363
Db 868 ACCGATATGAGACAGATGATGATACCAAGTAAAGCGGCTTGAAGAAGGGGACAAA 927
Qy 364 ValValaIleGlnIleSerIleAlaGlyIleThrGlyAlaLysIleValThrProLys 383
Db 928 GTGTCATCTCCCAATTAACCGCGCGA-GCAACAGAGAAAGCGGCAACCGCCCTACG 986
Qy 384 GlnTrpAlaSerSerGlnGlnGlnAlaAlaPro 395
Db 987 -----CGGCCGCGCGCG 1001

RESULT 38
AA253632
ID AA253632 standard; DNA; 1005 BP.
AC AA253632;
XX 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 290 partial DNA sequence SEQ ID NO:1213.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX anti-bacterial; gene therapy; ds.
OS Neisseria meningitidis.
XX
XX MO9957280-A2.
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99MO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galicotti C, Grandi G, Hickey E, Masigiani V, Mora M,
XX Petersen J, Pizsa M, Rapunoli R, Ratti G, Scalato E, Scarselli M,
XX Tettelein H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX P-PSDB; AAY74870.
XX
XX Novel Neisseria polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX
XX Claim 7; Page 676-677; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to

```

CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 1005 BP, 271 A, 233 C, 315 G, 186 T, 0 other;

Alignment Scores:

Pred. No.:	2,236-13	Length:	1005
Score:	271.00	Matches:	100
Percent Similarity:	45.45%	Conservative:	60
Best Local Similarity:	28.41%	Mismatches:	141
Query Match:	13.42%	Indels:	52
		Gaps:	10

US-09-889-756a-2 (1-412) x AAZ53632 (1-1005)

```

Qy 75 ValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlySerTyrVal 94
Db 7 GTCCGCGCGCAGCATCCGCGCGCAGATTAAGATCTTATCTCAAACTCGCGCAACAGGTT 66
Qy 95 ArgAlaGlyGlnProLeuTyrGlnIleAspSer-----SerThr 107
Db 67 AAAAAGGCGATTGATTCGGAATCAATTCTGACCTCGACACCAATACCTCAATACG 126
Qy 108 TyrGlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAla 127
Db 127 GAAAAATCCAGTTGGAACGTAATCAGCGCAAGCTGCTGCGCAGATTCATTGGGC 186
Qy 128 LysAlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArg 147
Db 187 AGCCGCGGAGAAATAATATATACGTCAGCGCGCTTATGAGAAAGCAAGCATTCCTCAAA 246
Qy 148 GlnGlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAla 167
Db 247 GAGATTTTGAAGCGCGCAGATGCTTTGCCCGCGCAAAATGTTGCCGACGTCG 306
Qy 168 GlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg----- 180
Db 307 AAGCTTTAATCAGACAGCAAGCAAAATTCCTCATATACCGCGCAGTGGGAATTGGGCTAC 366
Qy 181 SerArgIleThrAlaProIleSerGlyPheIleGlnSerLysValSerGlnGlyThr 200
Db 367 ACGGCGATTACCGCAACGATGACGCGCGTGGCGGATTCGTCGTAAGAGAGGGGACG 426
Qy 201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgIlnThrAspProMetTyr 220
Db 427 ACTGTGAACGCGCGC-----CAGTTCACGCCGACGAT 459
Qy 221 ValAsnValThrGlnSerAlaSerGlnValMetLysLeuArgArgGlnIleAlaGlnGly 240
Db 460 GTCCAAATGGCGAAT-----CTGATATGATGTTGAACAAATATGACATGTCGAGGGC 513
Qy 241 LysLeu-----LeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspGly 258
Db 514 GATATTCACCAAGGGAAGCGCGGCGAGATATTCGTTACGATTTGTCGGAACCGGAT 573
Qy 259 ThrValTyrProGlnLysGlyArgLeuLeuPheAlaAspProValVal----- 274
Db 574 ACG-----CCGATTAAAGCGCAACCTCGACGCGTCGACCCGCGGCTGACCAAGATGCG 627
Qy 275 -----AangLysThrGlyGlnIleThr-----LeuArg 284
Db 628 TCGGCGCGTTCAACACAGATACGATACGCGCTTCATACGCGGTCTACTATTATGCGCGT 687
Qy 285 AlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMet 304
Db 688 TCGTTTGTGCGGAATTCGCGAAGCAACTCGCCACAGGGAGATGACGACGCAAAATACGTT 747
Qy 305 AspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal---ThrArgGly 323
Db 748 GAAATCGACGCGCGTGAATAATGTGCTGATATTCGTCGTCGACCGTGAATAAATTCGCGCG 807
  
```

```

Qy 324 AlaLysAspThrValMetIleValAsnAlaGlnGlyMetGluProArgGluValThr 343
Db 808 GCGAAGCGCTTTGGCCGCTTGGTCGCGGACGCGCAAGCGCGCAACCGCAAAATCCG 867
Qy 344 ValAlaGlnGlnGlnGlyThrAsnTyrIleValThrSerGlyLeuLysAspGlyAspLys 363
Db 868 ACCCGTATGAGACAGCACTATGAATACCGAAGTAAAGCCGCTTGAAGAGCGGACAAA 927
Qy 364 ValValAlaGlnGlyIleSerIleAlaGlyIleThrGlyAlaLysLysValThrProLys 383
Db 928 GTGTCATCTCCGAAATTAACCGCCGCGA-GCAACAGAAAGCGCGCAACGCGCCCTTAG 986
Qy 384 GluTyrAlaSerSerGluAsnGlnAlaAlaAlaPro 395
Db 987 -----CGGCGCGCGCGCGCG 1001
  
```

RESULT 39

ABQ91605 standard; DNA, 900 BP.

ABQ91605:

01-OCT-2002 (first entry)

M. capsulatus gene #1590 for DNA array.

Micro array; gene; ds; differential expression; gene expression.

Methylococcus capsulatus.

WO200255655-A2.

18-JUL-2002.

14-JAN-2002; 2002WO-NO00019.

12-JAN-2001; 2001NO-0000235.

PR 12-JAN-2001; 2001NO-0000239.

XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.

PA (TIGR-) TIGR.

XX BitKeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;

XX Lillhaug JR, Lossius I, Eissen JA, Fraser CM, Durkin AS;

XX Salzberg SL;

XX WPI; 2002-557818/59.

PS Novel DNA array useful for determining differential expression of

PT Methylococcus capsulatus genes, comprises polynucleotides or

PT oligonucleotides representative for a selective number of Methylococcus

XX capsulatus genes -

XX Claim 14; Page 607; 678pp; English.

PS The invention relates to a novel DNA array giving a representation of a

CC number of Methylococcus capsulatus genes. The method of the invention is

CC useful for determination of the differential expression of the genes of

CC M. capsulatus, and for studying gene expression on a genomic scale and in

CC gene expression assays of M. capsulatus genes. The sequences shown in

CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the

Alignment Scores:

Pred. No.:	7,79e-11	Length:	900
Score:	239.50	Matches:	97
Percent Similarity:	47.71%	Conservative:	49
Best Local Similarity:	31.70%	Mismatches:	128
Query Match:	11.86%	Indels:	33
		Gaps:	10

[illegible]

XX	12-JUN-2002	(first entry)
DB	Oligonucleotide for detecting cytosine methylation SEQ ID NO 11553.	
XX		
XX	Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;	
XX	drug; side effect; cancer; central nervous system; cardiovascular;	
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;	
XX	SNP; cell differentiation; ds.	
OS	Homo sapiens.	
XX		
PN	WO200218632-A2.	
XX		
PD	07-MAR-2002.	
XX		
PF	01-SEP-2001; 2001WO-EP10074.	
XX		
FR	01-SEP-2000; 2000DE-1043826.	
PR	05-SEP-2000; 2000DE-1044543.	
XX		
PA	(EPiG-) EPIGENOMICS AG.	
XX		
PI	Olek A, Piepenbrock C, Berlin K, Guetig D;	
XX		
DR	WPI; 2002-371829/40.	
PT	Determining the degree of cytosine methylation in genomic DNA, useful	
PT	for diagnosis and prognosis, comprises selective hybridization of	
PT	amplicons from chemically treated DNA	
DS	Claim 12; 56pp + Sequence Listing; 56pp; German.	

	CC	This invention describes a novel method for determining the degree of
	CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a
	CC	genomic sample of DNA. The sample is treated chemically to convert
	CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
	CC	DNA that contains the target C is amplified to form a labeled amplicon.
	CC	The amplicon is hybridised to two peptides, each with at least one
	CC	C member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
	CC	and the degree of hybridisation to both classes is determined from the
	CC	label on the amplicon. From the ratio of labels hybridised to the two
	CC	classes of oligomers, the degree of methylation is calculated. The method
	CC	is used: (i) for diagnosis and/or prognosis of side effects of
	CC	therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
	CC	of the central nervous, cardiovascular, gastrointestinal and respiratory
	CC	systems etc., particularly by detecting mutations or single nucleotide
	CC	polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
	CC	types and for investigating cell differentiation. The method allows the
	CC	methylation status of many C residues to be determined simultaneously.
	CC	ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
	CC	method for determining the degree of cytosine methylation described in
	CC	the disclosure of the invention.
	XX	
SX	Sequence	607 BP; 151 A; 53 C; 176 G; 227 T; 0 other;
Alignment Scores:		
Pred. No.:	1.97e-09	Length: 607
Score:	220.00	Matches: 60
Percent Similarity:	53.51%	Conservative: 39
Best local Similarity:	32.43%	Mismatches: 84
Query Match:	10.90%	Indels: 2
DB:	24	Gaps: 1
US-09-889-756A-2 (1-412) x ABQ24962 (1-607)		
OY	187	IIESeRGlyPhneIIeGLyGlnSerLysValaSerGluGlyThrIleuLeuaAnaIagLYaSP 206
		: : : : : : :
Db	7	ATTAGCGGTCGATTTGGTAACTCAACGTGCAGAACGCCGATTTGGATGAACGGTTAG 66
		: : : : : : :
OY	207	ThrThrValleuAlaThrIleArgGlnThrAsnProMetLeuValaenValThrGlnSer 226
		: : : : : : :
Db	67	GGCATTCGGTGTAATCGTAGTACTAATTGATTCGATTTACGTTGATGATGATTTAGCTTT 126


```
OY 227 AlaSerGluValMetLysLeuArgGlnIleAlaGlnGlyLysLeuAlaAlaAsp 246
      ::::::::::::::::::::
Db 127 AGTAACGATTTTTCGTTGTAATAGCAATTGGCAATGTAAGTGAATAAGAGAAC 186
      ::::::::::::::::::::
OY 247 GlyValIleAlaValGlyIleLysPheAspGlyThrValTyrProGlnLysGlyArg 266
      ::::::::::::::::::::
Db 187 GGTAAAGTTAAAGTGTATGATTATTAAGTACGCGTATTAAGTTTCGTAGGACGCTACG 246
      ::::::::::::::::::::
OY 267 LeuLeuPheAlaAspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAla 286
      ::::::::::::::::::::
Db 247 TTGCAATTTTTCGACGTTATCGTTGATTAGATTATTTGGGTTTATTATTACGCGTTATT 306
      ::::::::::::::::::::
OY 287 ValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrValArgValIleuMetAspGln 306
      ::::::::::::::::::::
Db 307 TTTTCGATTCGCGATTATTTGTTGTCGGGTATGTCGCGCGTACGTTTGAAGAA 366
      ::::::::::::::::::::
OY 307 ValAlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp 326
      ::::::::::::::::::::
Db 367 GGGTTTAATTTAAAGTTATTTTAACTTCGTAATAGGCGGTAAATTCGTACGTCGCGTGC 426
      ::::::::::::::::::::
OY 327 -----ThrValMetIleValAsnAlaGlnGlyMetGluProArgGluValThrVal 344
      ::::::::::::::::::::
Db 427 GATGTTATCGTATTCGTATGTTGGCCGCGATGATAAGTGAATTCGTCGATCGTTGTA 486
      ::::::::::::::::::::
OY 345 AlaGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysVal 364
      ::::::::::::::::::::
Db 487 AGTTAGGTTATTTGGCGATAGTGGTGTGATAGAAAGTTGAAGTAGGCGATCGCGCTA 546
      ::::::::::::::::::::
OY 365 ValValGlnGlyIle 369
      ::::::::::::::::::::
Db 547 GTAATAGTGGGTTG 561
      ::::::::::::::::::::
```

Search completed: September 8, 2003, 06:03:16
Job time : 3427 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 8, 2003, 04:58:37 ; Search time 4846 Seconds
(without alignments)
3478.078 Million cell updates/sec

Title: US-09-889-756A-2
Sequence: 1 MAFYAFKAMRAAALAAVAL.....AAPQGVQTAASEATASEAE 412

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgm2.1/USPTO.spool/US09889756/runat.05092003.124324.23279/app.query.fasta.1.583
-DB=GenEmbl -QMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=Dits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09889756@CGN 1.1.3508 @runat.05092003.124324.23279 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_on :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_da :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_or :
21: em_ov :
22: em_ov :
23: em_ov :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :

29: em_vi :
30: em_htg_hum :
31: em_htg_inv :
32: em_htg_other :
33: em_htg_mus :
34: em_htg_pln :
35: em_htg_rod :
36: em_htg_mam :
37: em_htg_vrt :
38: em_sy :
39: em_htgo_hum :
40: em_htgo_mus :
41: em_htgo_other :

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2019	100.0	1239	6 AX027848	AX027848 Sequence
2	1993	98.7	326301	1 NMA622491	AL162757 Neisseria
3	1989	98.5	1239	6 AX391542	AX391542 Sequence
4	1989	98.5	14056	1 AE002521	AE002521 Neisseria
5	1989	98.5	349980	6 AX044033	AX044033 Sequence
6	1938	96.0	1500	1 NGU14993	U14993 Neisseria
7	1908.5	94.5	1235	1 AF037041	AF037041 Neisseria
8	1278	63.3	2127	1 NGMTRC	225796 N.gonorrhoea
9	839	41.6	215050	1 AL646057	AL646057 Ralstonia
10	777	38.5	20382	1 AE008856	AE008856 Salmomoni
11	770	38.1	11110	1 AE005757	AE005757 Callobact
12	765	37.9	13945	1 AE012380	AE012380 Xanthomon
13	763.5	37.8	6066	1 EAE306389	AJ306389 Enterobac
14	762.5	37.8	4879	6 E27981	M94248 E. coli acr
15	762.5	37.8	6039	1 ECUB0734	U00734 Escherichia
16	762.5	37.8	12524	1 AE005225	AE005225 Escherich
17	762.5	37.8	13446	1 AE000152	AE000152 Escherich
18	762.5	37.8	13987	1 AE000152	US2664 Escherichia
19	762.5	37.8	13987	1 AE000152	AE000152 Escherich
20	762.5	37.8	13987	1 AE000152	AE000152 Escherich
21	760.5	37.8	295741	1 AE016756	AE016756 Escherich
22	759.5	37.6	11927	1 AE015072	AE015072 Shigella
23	759.5	37.6	292309	1 AE016979	AE016979 Shigella
24	758.5	37.6	5281	1 AY061647	AY061647 Proteus m
25	758	37.5	197050	1 AL646081	AL646081 Ralstonia
26	756	37.4	240050	1 AL627267	AL627267 Salmomoni
27	756	37.4	300029	1 AE016842	AE016842 Salmomoni
28	753	37.3	5600	1 PSEENVCD	U11616 Pseudomonas
29	753	37.3	13987	1 AE004479	AE004479 Pseudomon
30	753	37.3	20347	1 AE008717	AE008717 Salmomoni
31	750.5	37.2	6803	1 ECEVNC	M96848 E. coli acr
32	750.5	37.2	11095	1 AE000405	X57948 E. coli envC
33	750.5	37.2	11095	1 AE000405	AE000405 Escherich
34	750.5	37.2	110000	1 ECOM671	Continuation (2 of
35	749.5	37.1	11807	1 AE005554	AE005554 Escherich
36	749.5	37.1	307962	1 AP002564	AP002564 Escherich
37	747	37.0	5914	1 SMA252200	AP002564 Escherich
38	745.5	36.9	300359	1 AE016767	AJ252200 Stenotrop
39	741.5	36.7	12478	1 AE011925	AE011925 Escherich
40	738	36.6	4503	1 AB104882	AB104882 Xanthomon
41	735.5	36.4	301995	1 AE016779	AE016779 Pseudomon
42	734	36.4	11048	1 AE015901	AE015901 Shewanell
43	732	36.3	110000	2 AC141230_0	AC141230 Homo sapi
44	731	36.2	254677	2 AC068494	AC068494 Mus muscu
45	730.5	36.2	12296	1 AF031417	AF031417 Pseudomon

RESULT 1

ALIGNMENTS

AX027848
 LOCUS AX027848 1239 bp DNA linear PAT 24-NOV-2000
 DEFINITION Sequence 1 from Patient M00043517.
 ACCESSION AX027848 AX027852
 VERSION AX027848.1 GI:10188692
 KEYWORDS
 ORGANISM *Neisseria meningitidis*
 SOURCE *Neisseria meningitidis*
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE
 1. Thomson, J.
 AUTHORs Babs055 polynucleotide and polypeptide from *neisseria meningitidis*.
 TITLE Uses thereof
 JOURNAL Patent: WO 0043517-A 1 27-JUL-2000;
 COMMENT SMITHKLINE BEECHAM BIOLOG (BE) ; THOMARD JOELLE (BE)
 FEATURES On Oct 15, 2002 this sequence version replaced gi:10188695.
 source Location/Qualifiers
 1. 1239
 /organism="Neisseria meningitidis"
 /mol_type="genomic DNA"
 /db_xref="taxon:487"
 BASE COUNT 275 a 332 c 384 g 248 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,21e-112 Length: 1239
 Score: 2019.00 Matches: 412
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 DB: 6
 US-09-889-756a-2 (1-412) x AX027848 (1-1239)

541 TCGGCATTACCGCGCCGATTTCCGGCTTATCGTCAGTCCAAAGTTTCCGAAGTACG 600
 Db
 201 LeuLeuSnaIaGlaGlaPThrThrValLeuAlaThrIleArgIlnTrpAspProMetTyr 220
 Oy
 601 CTGTTGAATGGGGCGATACGACCGTCTGGCAACCATCCGCCAAACCATCCGATGAT 660
 Db
 221 ValAsnValThrgInSerAlaSerGluValMetLeuLeuArgArgIlnIleGluGly 240
 Oy
 661 GTGAACCTTACCCAGTCTGCATCCGAAGTATGAAATTGGCCGCTCAGATAGCGAAGGC 720
 Db
 241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
 Oy
 721 AAACGTGTGGGGGGGATGGTGAATTCGGTCGCAATTCGAACGACGACAGT 780
 Db
 261 TyrProGluLysGlyArgLeuLeuPheAlaAspProValValAsnGluSerThrGlyGln 280
 Oy
 781 TACCTTAAAAAGGCCCTCTGCTCTTTGCCGATCCGCTCGTCAACGATGACCGGTACG 840
 Db
 281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
 Oy
 841 ATTACCTGGCGCGCGCGCGTACCGAACGATCAGATATCCGTGATGCGCGGTGTATGTG 900
 Db
 301 ArgValIleuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320
 Oy
 901 CGCGTGTGATGACCAAGTGGCGGTGATTAACCATTTGTGTGCGCGACGACGCGTGA 960
 Db
 321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArg 340
 Oy
 961 ACCGCGCGTGGCAAGATACCGTATATGTGATATCCCAAGCGGTATGGAACCCCGC 1020
 Db
 341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
 Oy
 1021 GAGGTAAAGGTTGGCAACAGCAGGATACGATGATGATGATGATGATGATGATGATGAT 1080
 Db
 361 GlyAspLysValValValGluGlyIleSerIleLeuAlaGlyIleThrGlyAlaLysVal 380
 Oy
 1081 GGGGACAAAGGTGTGTGGAAAGCATCATGATCCCGGTATACGGGTGCAAAAAGGTA 1140
 Db
 381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln 400
 Oy
 1141 ACCGCCAAGATGGGGGTGCTCTGAAACCAACGCCGCCGCCCTCATTCGGGCTTCAG 1200
 Db
 401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGln 412
 Oy
 1201 ACCGCATCTGAAGCCAAACGCTTTCGAAGCGAA 1236
 Db
 RESULT 2
 NMA622491/c 326301 bp DNA linear BCT 02-SEP-2002
 LOCUS NMA622491
 DEFINITION *Neisseria meningitidis* serogroup A strain Z2491 complete genome;
 segment 6/7.
 ACCESSION AL162757 AL157959
 VERSION AL162757.2 GI:7380371
 KEYWORDS
 ORGANISM *Neisseria meningitidis* Z2491
 SOURCE *Neisseria meningitidis* Z2491
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE
 1 (bases 1 to 326301)
 AUTHORs Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C.,
 Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T.,
 Davies, R. M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N.,
 Holtroyd, S., Jagels, K., Leather, S., Moule, S., Mungall, K.,
 Quail, M. A., Rajandream, M. A., Rutherford, K. M., Simmonds, M.,
 Skellern, J., Whitehead, S., Spratt, B. G. and Barrall, B. G.
 TITLE Complete DNA sequence of a serogroup A strain of *Neisseria*
 meningitidis Z2491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE
 2 (bases 1 to 326301)
 AUTHORs Parkhill, J.
 TITLE Direct Submission

JOURNAL

COMMENT

Submitted (30-MAR-2000) Submitted on behalf of the Neisseria Sequencing Team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:
Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES

source

1. 326301
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
196. 311
/note="REP 2; hmms hit to HM REP 2 (1 - 109), score: 87.43"
434. 453
/note=">= 90% match to ATTCCNNNNNNNGGAAT"
/label=dR63
498. 560
/note="Correia element; hmms hit to HM Correia (1 - 62), score: 78.71"
/label=Correia
562. 604
/note="Correia element; hmms hit to HM Correia (114 - 156), score: 51.92"
/label=Correia
complement(667. 676)
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS
complement(668. 1813)
/gene="daps"
complement(668. 1813)
/gene="daps"
/BC_number="3.5.1.18"
/note="NMA1730, daps, succinyl-diaminopimelate desuccinylase, len: 381aa; similar to many eg. SW:P24176 (DAP6 ECO1) daps, succinyl-diaminopimelate desuccinylase from Escherichia coli (375 aa) fasta scores: E(): 0.56, 9% identity in 371 aa overlap. Contains Pfam match to entry PF01546 Peptidase_M20."
/codon_start=1
/transl_table=11
/product="succinyl-diaminopimelate desuccinylase"
/protein_id="CAB84958.1"
/db_xref="GI:7380372"
/db_xref="SPTRMBL:Q9JTL0"
/translation="MTETSLLEAKLISRPVTPDDRCOKLVERLYKIGFAEELHFGDTKNIWLRGKVPVCFAGHTDVVPFGPEKMSPPPEPTERDRLYRGAAVMKTSIACPVTAECERVAEPHOGSIALILTSDEGALDGTTRVVDLTKARGLIYCVIGSEPTAVDVLGDMIKNGRGSLSGNTVKGKCHAIYPLAINTPAPALLEITQEVNDEGKATFPPTSPQISNINSGIGATNVPGLNVPKFNFRSTESTEALKORVAHILDRHGVYDLQWCSGQPLTQAGKLTVDVARAAIETCGIEALSTTGSGISDRFKAIABELIEGSAITIHQINENVRLDIPKLISAVEGILLALAGNAV"
complement(836. 1786)
/gene="daps"
/note="Pfam match to entry PF01546 Peptidase_M20, score 252.50, E-value 5.7e-72"
252.50, E-value 5.7e-72
complement(1863. 1872)
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS
complement(1927. 2547)
/gene="NMA1731"
complement(1927. 2547)
/gene="NMA1731"
/note="NMA1731, conserved hypothetical protein, len: 206aa; similar to SW:P44013 (Y552 HAEIN) hypothetical protein from Haemophilus influenzae (207 aa) fasta scores: E(): 0.59, 6% identity in 203 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

RBS

/protein_id="CAB84959.1"
/db_xref="GI:7380373"
/db_xref="SPTRMBL:Q9JQ23"
/translation="MLTRKSCDLFNIPFQFQSLKKYQPSIPQIKADYKNNQIMQO LIOOVAELGAPAPAPRIEBCWGCWQRAFPAYFKYKQKNSAIIISLNRRLSV SLIMHCYKADVSPTALPDINRMIDNFDTEKTSADDMWGHSEYDTRTYAQQNESDR KLNQDEDFPCIGHIERDDIGRQDVAKWIAETVEDLLPLYEACHGK"
complement(2551. 2558)
complement(2570. 2579)
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS
complement(2581. 3090)
/gene="NMA1732"
complement(2581. 3090)
/gene="NMA1732"
/note="NMA1732, conserved hypothetical protein, len: 169aa; similar to SW:P43971 (Y243 HAEIN) hypothetical protein from Haemophilus influenzae (172 aa) fasta scores: E(): 0.49, 7% identity in 169 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB84960.1"
/db_xref="GI:7380374"
/db_xref="SPTRMBL:Q9JTK9"
/translation="NNPFQSVTPAEPLEMLYACHGVKVRPCGOVAMLSDYLAENG NOLVLOTIOITVFNVAAPLHDEEENFPFLLOYAPOASVDELLROHYSIHN WDAVAEPKLENDNVIYIDAEFKFVACIVDHALIEPLFMGKKFIPEKLTIEG EIVAKRRR"
complement(3185. 3736)
/gene="NMA1733"
complement(3185. 3736)
/gene="NMA1733"
/note="NMA1733, H.8 outer membrane protein, len: 183aa; strongly similar to SW:P07212 (H8_NIME) H.8 outer membrane protein from Neisseria meningitidis (183 aa) fasta scores: E(): 0.96, 7% identity in 183 aa overlap. Contains Pfam match to entry PF00127 copper-bind, Copper binding proteins, plastocyanin/azurin family and, Prosite match to PS00196 Type-1 copper (blue) proteins signature."
/codon_start=1
/transl_table=11
/product="H.8 outer membrane protein"
/protein_id="CAB84961.1"
/db_xref="GI:7380375"
/db_xref="SWISS-PROT:P57025"
/translation="MKAYLATISAIVIGLAACQEPAPAEATPPAAEPASAPAE AAPADAAEPAGNCATVENSNDMMQFNKDIQVSAKCEFTILLHGTGTPPTSGH NIVIGTEBMDGIFKDGVAADTDYKPPDARVVAHTKLIIGGEESLTLDPKADG EYKFACTFPGHGLNMKGVTIWD"
complement(3191. 3568)
/gene="NMA1733"
/note="Pfam match to entry PF00127 copper-bind, Copper binding proteins, plastocyanin/azurin family, score 218.60, E-value 9.1e-62"
218.60, E-value 9.1e-62
complement(3212. 3259)
/gene="NMA1733"
/note="PS00496 Type-1 copper (blue) proteins signature."
complement(4017. 4439)
/gene="NMA1734"
complement(4017. 4439)
/gene="NMA1734"
/note="NMA1734, unknown, len: 140aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB84962.1"
/db_xref="GI:7380376"
/db_xref="SPTRMBL:Q9JTK6"
/translation="NMLTGRRDALSLGLKQAGWRRLDAAYKKLLPANIHPHFO TACTEDGRLVLAANNMAASRLKMLAPSVLPQLAGDASIRSVSVRLVPEKPEPKNT IHLISVALESFQSAVVKLEKRHPLEALANLVRHGA"
4306. 4315

misc_feature

complement(667. 676)
/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS
complement(668. 1813)
/gene="daps"
complement(668. 1813)
/gene="daps"

gene

complement(668. 1813)
/gene="daps"

CDS

complement(668. 1813)
/gene="daps"

misc_feature

complement(3191. 3568)
/gene="NMA1733"

misc_feature

complement(3212. 3259)
/gene="NMA1733"

misc_feature

complement(1863. 1872)
/note="Core DNA uptake sequence: gccgcctcgaa"

gene

complement(1927. 2547)
/gene="NMA1731"

CDS

complement(1927. 2547)
/gene="NMA1731"

misc_feature

complement(4017. 4439)
/gene="NMA1734"

```

/note="Core DNA uptake sequence: gccgccttgaag"
/label=DUS
/complement(4486..4495)
/note="Core DNA uptake sequence: gccgcttcgaag"
/label=DUS
/size=7320
/gene="seca"
4570..7320
/note="secA"
/gene="WMA1735, seca, preprotein translocase SecA subunit."
len: 916aa; similar to many eg. SW:P10408 (SECA_ECOLI)
preprotein translocase SecA subunit from Escherichia coli
(901 aa) fasta scores; E(): 0.58.0% identity in 920 aa
overlap. Contains Pfam match to entry PF01043
SecA protein, SecA protein, amino terminal region and
Prosite match to PS01312 Protein secA signatures.."
/codon_start=1
/transl_table=1
/product="preprotein translocase SecA subunit"
/protein_id="CAB84963.1".
/db_xref="db:"GI:7380377"
/db_xref="SPTREMBL:Q9JTK7"
/translation="MLTNIIAKKI FGSNRDELIKOYKASVARIALIEEOMALSPADLIQLSDADLCLE
AKTAEFKRLADGGTLGILPEAPAVCRSEARSTTEGMRFVDVLIGVAGLDHGKIAELNEDGKIAELNED
RTGSKETLVATLAIVLAGKGVAITVDNVTVLASDADIIMELPYNFGLTGVIISL
MOPFERQAAVADITYGTNNRGEFDYIRDMVTYDDQYGRLEINRAVYDEVDSLLIDB
ARTPLIISGDADDNIOLYOTIMNTVPPLRVKOEFTSESQSRYWDEKAHOVILLSEAGEHE
AEQILTONGLEAGHQAWEKAEKMEIIRKEQTIASIFTPNYPRALTYKLSGMTGADTEI
IGRLRMGRMEEGHQAVEKAEKMEIIRKEQTIASIFTPNYPRALTYKLSGMTGADTEI

```

Alignment Scores:

Pred. No.:	1.67e-
Score:	1993.0
Percent Similarity:	99.03%
Best Local Similarity:	98.54%
Query Match:	98.71%
DB:	1

US-09-889-756A-2 (1-412) X NMA6Z2491 (1-326301)

Oy	1	MetalAlpHeTyrAlaPheLysAlaMetArgAlaAlaAlaAlaLeuAlaAlaAlaValAlaLeu	20
Db	245789	ATGGCTTTTAAAGCTTTTAAAGGGAAGCGTGGCGCGCTGGCTCCGCTTGGATTG	245730
Oy	21	ValLeuSerSerCysG61YgysG1ygiAspAlaAlaG1ng1ygiInProAlaIArg	40
Db	245729	GTACTGTGCTCTTGGCGGTAAAGCGGAGACGCGCGCACGGCGGAGAGCTGTGTCGG	245670
Oy	41	GIuaIaProAlaProValValGIyValValThrValHisProG1nThrValAlaLeuThr	60
Db	245669	GAAGGCCCTGGCCCGCTCGTGGTGCATGACGTCATCCGCAACCGTCGATTGACC	245610
Oy	61	ValGIuLeuProGIyYArgLeuG1uSerLeuArgThrAlaAspValArgAlaG1nValGIy	80
Db	245609	GTCGAGTTGCCGGGGCGTTTGGAAATGCTGTCGACCGCCGATTCGCCGCCCAAGTCGGC	245550
Oy	81	GIy11e11eG1nLysArgLeuPheG1ng1ySerTyrValArgAlaG1yG1nProLeu	100
Db	245549	GGCATCATCTCAAAAACGCGCTGTTCCAAAGAGGACAGTTAAGTCGCTCCGACACGGCGCTG	245490
Oy	101	TyrG1n1LeaSpSerSerThrTyrG1ua1aAsnLeuG1uSer1aArgAlaG1nLeuAla	120
Db	245489	TATCGATGACAGTTCCACTTATGAAAGCAGCTGTGGAAAGTCCGCCGCCCAACTGGCA	245430
Oy	121	ThrAlaG1nAlaThrLeuAlaLysAlaAspAlaAspLeuAlaArgTyrLysProLeuVal	140
Db	245429	ACGGCTCAGGCAACGCTTCCAAAGCGGATGCGGATTTGGCGCGGTACAAAGCTTTGGT	245370
Oy	141	AlaAlaG1ua1aValSerArgG1ng1yTyrAspAlaAlaValThrAlaLysArgSerAla	160
Db	245369	GCGCGCGAAGCGCTACGCGCGGAGAAATACGATGCTGCGGTAAACGGCGAATCCTTGGCC	245310
Oy	161	GIua1aG1yValLysAlaAlaG1nAlaAlaAlaLeuSerAlaG1y11eAsnLeuAsnArg	180

[illegible]

```

/nc="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD28957.1"
/db_xref="GI:19700150"
/translation="MAFAPFKMRAAALAAVALVSSCGKGDAAAGCGPAGREARA
PVGAVVHPQVLAIVELPGRLESLRTADVRAOVGIIQRLRFOESYVAGPLYO
IDSTYEGALSSARAOATLAQATLAKADADLARXPLVAAASVROEYDAVATKRSR
EAGYAAQAAIKSAGISLNRRTAPISGFIQSGKSEGTLLNAGDAVTLATIRQTP
MYNVTGASSEWMLRROJAEGLKLAADGVAVGIPDGVPEKGRLEPADPANE
STGQITLRAVNDONILPGLVYRVMDQVADNAPVPOQATRAKQTPVTVNQ
GGMREPVTAOQOQTNMTVTSGLKDGKRVVEGISTAGITGAKVTPKEMASENQA
AAPSGVGTASEAPASEAK"
BASE COUNT      269 a      337 c      388 g      245 t
ORIGIN
Alignment Scores:
Pred. No.:      1,99e-110      Length:      1239
Score:          1989.00      Matches:      406
Percent Similarity: 99.03%      Conservative: 2
Best Local Similarity: 98.54%      Mismatches: 4
Query Match:     98.51%      Indels:      0
DB:              Gaps:      0
US-09-889-756a-2 (1-412) x AX391542 (1-1239)
QY      1 MetAlaPheTYrAlaPheLYaAlaMeCArgAlaAlaAlaLeuAlaAlaValAlaLeu 20
Db      1 ATGGCTTTTAACTGTTTAAAGCGCATGCGTCCGCGCTTGCGCCGCTTGCACTTG 60
QY      21 ValLeuSerSerCYsGLyLYsGLyGLyAspAlaAlaGlnGLyGLyGlnProAlaGLyArg 40
Db      61 GTACTGTCGCTCTCGGTAAGCGAGACCGGCGGCGGCGGCGGCGGCGGCTGTCGCGG 120
QY      41 GluAlaProAlaProValValGLyValValThyValHisProGlnThrValAlaLeuThr 60
Db      121 GAAACCCCTCGCGCCGTCGTGCTGCTAACCTGCAACCGTCCGCACTTGAACC 180
QY      61 ValGluLeuProGLyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGLy 80
Db      181 GTCCAGTTGCCGCGCGCTTTGAATCGCTGCGTACCGCGCATGTCGCGCCCAAGTCGCG 240
QY      81 GlyIleIleGlnLYsArgLeuPheGlnGlnGLySerTYrValArgAlaGLyGlnProLeu 100
Db      241 GGCATCATCCAAAACGCTGTTCCAGAAGACGATTATGTCGTCGCGGACACCGCTG 300
QY      101 TYrGlnIleAspSerSerThrTYrGLuAlaAsnLeuGlnSerAlaArgAlaGlnLeuAla 120
Db      301 TATCAGATCGACAGTTCACATTATGAAGCAGCTCGAAGACGCGCGCGCAACTGGCA 360
QY      121 ThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeuAlaArgTYrLYsProLeuVal 140
Db      361 ACGGCTCAGGCAAGCTTGCCAAAGCGATGCGGATTTGGCGCATCAACGCTTTGGTT 420
QY      141 AlaAlaGlnAlaValSerArgGlnGLyTYrAspAlaAlaValThrAlaLYsArgSerAla 160
Db      421 GCGCGCCGAAGCGTCAGCGCGCAGGAATACGATCTCGGTAACGGGAAACGTTTCCGC 480
QY      161 GluAlaGLyValLYsAlaAlaGlnAlaAlaIleLYsSerAlaGLyIleAsnLeuAsnArg 180
Db      481 GAGCGACGCGTTAAAGCGCGCAGCGCATCAAAATCCGCGGCATCAGCTGAACCGT 540
QY      181 SerArgIleThrAlaProIleSerGLyPheIleGLyGlnSerLYsValSerGlnGLyThr 200
Db      541 TCGGCGATTACCGCGCGCATTTCCGCTTTATCCGTCATCCAAAGTTCCGAAGGTAAG 600
QY      201 LeuLeuAsnAlaGLyAspThrThrValLeuAlaThrIleArgGlnThrAsnProMetTYr 220
Db      601 TTGCTGAACGCTGGCGATGCAACCGTACTGCGCAGCAATCCGCCAAACCAATCCGATAT 660
QY      221 ValAsnValThrGlnSerAlaSerGluValMetLYsLeuArgArgGlnIleAlaGLy 240
Db      661 GTGAACGTTACCCAGTCTGCATCCGAAGTATGAAATTTGGCGCTCAGATAGCGGAAGGC 720

```

```

QY      241 LysLeuLeuAlaAlaAspGLyValIleAlaValGLyIleLYsPheAspAspGLyThrVal 260
Db      721 AAATCTGCGCGCGGATGGTGTATTCGGTGGCATCAAAATTTGACGACGACAGTT 780
QY      261 TYrProGlnLYsGLyArgLeuPheAlaAspProValValaGlnGlnSerThrGLyGln 280
Db      781 TACCTGAAAAGGCGCCCTGCTGTTGCCGATCCGCGCATCAAGATGACCGGTCAG 840
QY      281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGLyLeuTYrVal 300
Db      841 ATTACCTTGGCGCGCGCGATCCGAACGATCAGATATCTTGATGCGCGGCTGTATGTG 900
QY      301 ArgValLeuMetAspGlnValAlaValaAspAsnAlaPheValValProGlnGlnAlaVal 320
Db      901 CGCGTCTGATGACCAAGTGGCGGTGGATGATACCGCATTTGTTGGCCGACGACGGCGTA 960
QY      321 ThrArgGLyAlaLYsAspThrValMetIleValaAsnAlaGlnGLyMetGluProArg 340
Db      961 ACGCGCGGTGCGAAGATACCGTATGATGATGATCCCAAGCGGATGGAACCCCGC 1020
QY      341 GluValThrValAlaGlnGlnGlnGLyThrAsnThrIleValThrSerGLyLeuLYsAsp 360
Db      1021 GAGGTAAACGTTGGCGCAACAGAGGTGACGATTTGATGCTGCGGCTGGAAGCAC 1080
QY      361 GlyAspLYsValValValGlnGLyIleSerIleAlaGLyIleThrGlyAlaLYsVal 380
Db      1081 GGGGACAGAGTGGTGTGGAAGCATCATGATCCCGGTATTAACGGTGGCAAAAAGTA 1140
QY      381 ThrProLYsGlnThrAlaSerSerGluAsnGlnAlaAlaAspProGlnSerGLyValGln 400
Db      1141 ACGGCCAAAGATAGGCGCTGCTGGAACCAAGCCCGCGCTCATCCGCGGTTACG 1200
QY      401 ThrAlaSerGlnAlaLYsThrAlaSerGluAlaGln 412
Db      1201 ACGGCATCTGAAGCCAAACCTGCTTGAAGCGAAA 1236
RESULT 4
AE002521/c 14056 bp DNA linear BCT 25-MAY-2000
LOCUS      Neisseria meningitidis serogroup B strain MC58 section 163 of 206
DEFINITION of the complete genome.
ACCESSION AE002521 AE002098
VERSION   AE002521.1 GI:7226962
KEYWORDS  Neisseria meningitidis MC58
SOURCE    Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
ORGANISM  Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 14056)
AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
          Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
          Dodson,R.J., Nelson,W.C., Gwinn,M.L., Deboy,R., Peterson,J.D.,
          Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
          Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
          Clifton,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
          Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Maignani,V.,
          Rappelli,R., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
          Rapinoli,R. and Venter,J.C.
          Complete genome sequence of Neisseria meningitidis serogroup B
          strain MC58
          Science 287 (5459), 1809-1815 (2000)
JOURNAL   MEDLINE
PUBMED    20175755
10710307
2 (bases 1 to 14056)
REFERENCES Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
          Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
          Dodson,R.J., Nelson,W.C., Gwinn,M.L., Deboy,R., Peterson,J.D.,
          Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
          Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
          Clifton,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
          Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Maignani,V.,
          Pizze,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,

```

TITLE Rappunli, R. and Venter, J.C.
Journal Direct Submission
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Source Location/Qualifiers
1. .14056
/organism="Neisseria meningitidis MC58"
/mol_type="genomic DNA"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
120. .1457
/gene="NMB1707"
120. .1457
/gene="NMB1707"
/note="similar to GB:AL009126 percent identity: 75.06;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="sodium- and chloride-dependent transporter"
/protein_id="AAF42055.1"
/db_xref="GI:7226963"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

```
/translation="MSNNTSSSKIGFVLAAGSAIGLAIWKRPPTAGTNGAVFEL  
LFLFTILVLPVQAEFYIGRTGKNAVDSFVLRPEQTQMLWGRMGVAACTLLSP  
YSVVGWLVNIVVHSTGAVHTGADFEALFGATISNPGSLSYQALFMLITVWVKG  
ISDGIKXANRYIMFGLFILFALAIRSLTLPGMEGVSLKPNWSYFKADMTITAG  
OAPFALSTIGVAMITVASYLGKQDMPSGHTIMMNLVSLAGLVFPVAFGE  
PSGPGILFIVLPVPMKMPGCTGAFVFMILVYFATLTSFMLETVIATSTRDOR  
KRKHHTLIGTALITIGIPSLASFGWGEFVPGKTIPTDMDYVTSAVINIGLSS  
IFTAMIDKQSVLADAGSTVPRAVLMLMTLRVLAIAIIVINSIDL"  
/complement (2297. .3091)  
/gene="NMB1709"  
/complement (2297. .3091)  
/gene="NMB1709"  
/note="similar to PID:2258280 percent identity: 98.49;  
identified by sequence similarity; putative"  
/codon_start=1  
/transl_table=1  
/product="chymidylate synthase"  
/protein_id="AAF42056.1"  
/db_xref="GI:7226964"  
/translation="WKAYLDLMRHVLNDGTDKSDRTGTGTRSVFQYQMRFDLKGFP  
LTTKKLHRSIIHELMFELKGDITNIKYKANNVSIWEMADENDLGVYGYQWRNP  
APDGRHIDQIANVLEQIKNPDSRLIYSANNPALVDPMALPCGALFQVFAAGKLS  
COLYORSADIFLGVFNIAYSALLTMMMAVCGEAGEFVHTPGDHLVYRHHFQOAL  
OLEBRPRALPYMKINPEVKDLFSKFEDELEGGDPRHITKAIVS"  
3475. .4809  
/gene="NMB1710"  
3475. .4809  
/gene="NMB1710"  
/note="similar to GP:146126 percent identity: 80.73;  
identified by sequence similarity; putative"  
/codon_start=1  
/transl_table=1  
/product="glutamate dehydrogenase, NADP-specific"  
/protein_id="AAF42057.1"  
/db_xref="GI:7226965"  
/translation="WTDLNTLFAULKQNPNOEPHQAVEEFMSLDPEFLAKNKYTO  
QSLRIVPERVWFRVTWQDDKQOVNNGIRYOMSSAIGYIKGIRFHYDVLGY  
LKFLAEQVFNKALTTLPWGKGGSDDPKGSDEAVMRCQAFMELVYHIGADTD  
VPAGDIGVGREIIGYLFQYKRIINEFSVLTGKLEWGSGLIREATGYCVFAOA  
MLQTNDSFEGRVLISSGNVAOYAAEKALOLGAKELTVSDNSGFVLPDGGMTAA  
LAALIELEKRRERVAATAKGOLOVEPKOKPMGVAAEIALPCATONLDEAKKTL  
ANGCYVABGANPESTIGAVQFITAAGILYAPGRASNAGVATSGLENSQAIRLSST  
REBVDRFLFGIMQSIHESCLYKGVDTVNVNGANIAGFVKVADAMLAQGF"  
/complement (4944. .5723)  
/gene="NMB1711"  
/complement (4944. .5723)  
/gene="NMB1711"  
/note="similar to PID:455922 GB:U00096 SP:P06957  
PID:1786303 PID:434009 percent identity: 53.25; identified  
by sequence similarity; putative"  
/codon_start=1
```

gene
CDS

gene
CDS

gene
CDS

```
/transl_table=1  
/product="transcriptional regulator, GntR family"  
/protein_id="AAF42058.1"  
/db_xref="GI:7226966"  
/translation="MKLVPRQKISDOVLSYLERIAGVYAEAGCKIPREVAEEGV  
SRPSVSALNVLVARQVLEARGQGYTVSKPODFLOSQOBLGHSNMEQVFPDS  
CHLEGCGAALAAERITDADIKRIGFWLEKEEESGESSGNHQSSEAVSFPQTLADAH  
NLFPSHLGGLLKMLYQTRSSLIYNQEDPDPKMAQRRVLYEAI SNRRPGEASBA  
AKAHLNVYVASSIILKREYOSRRHADTLAQNDIKRQVDMV"  
6154. .6453  
/gene="NMB1712"  
6154. .6453  
/gene="NMB1712"  
/note="similar to PID:882504 GB:U00096 SP:Q46839  
PID:1789347 percent identity: 71.01; identified by  
sequence similarity; putative"  
/codon_start=1  
/transl_table=1  
/product="L-lactate permease-related protein"  
/protein_id="AAF42059.1"  
/db_xref="GI:7226967"  
/translation="MORCFMPPHYCRBLEKRYRSGSINMETWQNTAIGSLYLTAA  
AALLPYVFPALTYLKLKGYQGLTYLLIALAAVAFGRMPGMVAVSSLPQPH"  
/complement (6839. .7804)  
/gene="NMB1713"  
/complement (6839. .7804)  
/gene="NMB1713"  
/note="similar to SP:P37247 PID:1197006 percent identity:  
61.94; identified by sequence similarity; putative"  
/codon_start=1  
/transl_table=1  
/product="transposase, IS30 family"  
/protein_id="AAF42060.1"  
/db_xref="GI:7226968"  
/translation="WSTYQLTQGERHYIQLNSHCYTEAKLAKNKKSTISREIRH  
RTGQGYSAKAGQSRRTIKQRKQPKLDSQILQHDITIRKLSPEQVCAVLCXH  
RITLHSTIYRVLRQDSNGSLTMOHLRISCKEYRRYGTWGRKVPNVGLENRA  
IVQKSRITGMEADTIVGKQKSLTLVETVRYTILCQLDLKADETARAVALAK  
AHDQVHTITMDNGKEFYQHTKITKALKATYVCVPHYSKEGLNENTNGLIRQYFPK  
QTFPNTSDEIRVQDELNHRPRKTLGYTPEVFLNLQPLIH"  
/complement (7914. .9317)  
/gene="NMB1714"  
/complement (7914. .9317)  
/gene="NMB1714"  
/note="similar to GB:L23839 PID:404773 PID:833173 percent  
identity: 64.15; identified by sequence similarity;  
putative"  
/codon_start=1  
/transl_table=1  
/product="multidrug efflux pump channel protein"  
/protein_id="AAF42061.1"  
/db_xref="GI:7226969"  
/translation="WDTTLKTYLTSVAAAFALSACTMIPYEQPKVEVATFRKDTAD  
SGIRAVDLGWHDFPADRLQKLDIALERTSLRTAVLNSEIRKQWIERNNLLPL  
AAVANDRSQSLSGGVSSYKVLGAASVELDFGVRSSEKALQGYFAVARRBA  
AHLSLIATVAKAVENERVYAEAMSIAQVSTKTEFKSELYKAGVIAVALLROE  
ALTESAKADVAHARSROARNALATLINQPIREDPAGLPLQDFVEXLPGISSE  
VLDRPPIRAEAHAKQANMIGAAAPAFPSIRLTCTGTGSAEIGCLFKSGGWS  
FASITLPIFTWGNKANDVAKLRQOVQVAVESAVQSFQVYANALARBQDLKAY  
DALSKGRARKEALRLVGLRYKGVSGALDLLDAERSVYABGAALSAQLTRANELAD  
LYAKGGLRQEDTQDK"  
/complement (9372. .12575)  
/gene="NMB1715"  
/complement (9372. .12575)  
/gene="NMB1715"  
/note="similar to PID:1408202 percent identity: 98.59;  
identified by sequence similarity; putative"  
/codon_start=1  
/transl_table=1  
/product="multiple transferable resistance system protein  
MtrD"  
/protein_id="AAF42062.1"  
/db_xref="GI:7226970"
```


ORIGIN

Alignment Scores:

Pred. No.:	3,18e-107	Length:	349980
Score:	1989.00	Matches:	406
Percent Similarity:	99.03%	Conservative:	2
Best Local Similarity:	98.54%	Mismatches:	4
Query Match:	98.51%	Indels:	0
		Gaps:	0

US-09-889-756a-2 (1-412) x AX044033 (1-349980)

```

QY      1 MetAlaPheTyrAlaPheLysAlaMetArgAlaAlaAlaAlaAlaAlaValAlaLeu 20
Db      296382 ATGGCTTTTATGCTTTTAAAGCCGATGCGTGGCGCGCGCTGGCTGGCGCGCTGGCATTTG 296323

QY      21 ValLeuSerSerCyGlyLysGlyLysAlaAlaAlaGlnGlyGlnProAlaGlyArg 40
Db      296322 GTACTGTCGCTTGGGTTAAAGCGGAGACGCGGCGGAGGCGGCGGCGGCGGCTGCTGG 296263

QY      41 GluAlaProAlaProAlaValAlaGlyValAlaThrValHisProGlnThrValAlaLeuThr 60
Db      296262 GAAGCCCTCGCGCGCGCTGCGTGTGCTTAACCGTCCATCCGCAAAACCGTGCATTGACC 296203

QY      61 ValGluLeuProGlyArgLeuGluSerLeuArgThrAlaAspValArgAlaGlnValGly 80
Db      296202 GTCGAGTTGCGCGGCGGCTTTGGAATCGCTGCGTACCGCGCATGTCGCGCCCAAGTCGCGC 296143

QY      81 GlyIleIleGlnLysArgLeuPheGlnGlnGlySerTyrValArgAlaGlyGlnProLeu 100
Db      296142 GGCATCATCCAAAACGCTGTCCAGAAAGCGAGTATATGTCGCGCGGAGCAGCGCTG 296083

QY      101 TyrGlnIleAspSerSerThrTyrGlnAlaLeuGlnSerAlaArgAlaGlnLeuAla 120
Db      296082 TATCGATCGACAGTTCCACTTATACAGCAGCTCGAAAGCGCGCGCGCACTGCGCA 296023

QY      121 ThrAlaGlnAlaThrLeuAlaLysAlaAspLeuAlaArgTyrLysProLeuVal 140
Db      296022 ACGGCTGAGGACGCTTCCAAAGCGGATGGGATTTGGCGCATACAAAGCTTTGTT 295963

QY      141 AlaAlaGlnAlaValSerArgGlnGlyTyrAspAlaAlaValThrAlaLysArgSerAla 160
Db      295962 GCCCGCGAAGCGCTGACGCGGAGAGAAATACATGCTGCGGTAAACGCGCAAAAGTTCTGCC 295903

QY      161 GluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSerAlaGlyIleAsnLeuAsnArg 180
Db      295902 GAGGACGCGCTTAAAGCGCGGACGCGGCAATCAATCCGCGCATGACCTGAACCT 295843

QY      181 SerArgIleThrAlaProIleSerGlyPheIleGlyGlnSerLysValSerGlnGlyThr 200
Db      295842 TCGGCGATTACCGCGCGCATTTCCGCTTTATCGGTCAGTCCAAAGTTTCCGAAGGTACG 295783

QY      201 LeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIleArgGlnThrAspProMetTyr 220
Db      295782 TTGCTGAACGCTGGGATGCGACACGCTACTGCGACCATCCGCCAAACCAATCCGATGAT 295723

QY      221 ValAsnValThrGlnSerAlaSerGlyValMetLysLeuArgArgGlnIleAlaGlnGly 240
Db      295722 GTGAACGTTATCCCACTGCTGCATCCGAAGTGATGAATGGCGCGCTGCATAGCCGAAGC 295663

QY      241 LysLeuLeuAlaAlaAspGlyValIleAlaValGlyIleLysPheAspAspGlyThrVal 260
Db      295662 AAACGCTGGCGGCGGATGCTGTGATTCGCGTCCGATCAATTTGACGACGCGCACAGTT 295603

QY      261 TyrProGlnLysGlyArgLeuLeuPheAlaAspProValValAsnGlnSerThrGlyGln 280
Db      295602 TACCTTGAAAAAGCGCGCTGCTGTTCGCCATCCGCGCGCATCAAGAAATCGACCGGTACG 295543

QY      281 IleThrLeuArgAlaAlaValProAsnAspGlnAsnIleLeuMetProGlyLeuTyrVal 300
Db      295542 ATTACCTGCGCGCGCGCGGACGACGATCAAGAAATATCTTGATGCGCGGCTGTATGTG 295483

QY      301 ArgValLeuMetAspGlnValAlaValAspAsnAlaPheValValProGlnGlnAlaVal 320

```

```

Db      295482 CGCGTCTGATGACCAAGTGGCGGTGATACGCACTTGTGTGGCCGACGACGCGTA 295423

QY      321 ThrArgGlyAlaLysAspThrValMetIleValAsnAlaGlnGlyMetGluProArg 340
Db      295422 ACGCGCGGTGGAAAGATACCGATGATTTGTGATGATCCCAAGCGGTATGAAACCCCGC 295363

QY      341 GluValThrValAlaGlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAsp 360
Db      295362 GAGGTACGGTTGGCAACGACGAGGTACGAATTGATTTGATGCTGGGTGTGAAGAC 295303

QY      361 GlyAspLysValValAlaGlnGlyLysSerIleAlaGlyIleThrGlyAlaLysVal 380
Db      295302 GGGGACAGGTGTGTGGAGAGCATCATGATCCCGGTATACCGGTGCGAAAGATA 295243

QY      381 ThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln 400
Db      295242 ACGCCCAAGATGGGGGTGCTGTGAAACCAAGCCCGCGCTTCAATCCGCGCTTACG 295183

QY      401 ThrAlaSerGluAlaLysThrAlaSerGluAlaGlu 412
Db      295182 ACGGCATCTGAAGCCAAACCGCTTCTGAAGCGAAA 295147

RESULT 6
NGU14993
LOCUS
DEFINITION
Neisseria gonorrhoeae membrane fusion protein (mtrc) gene, complete cds.
ACCESSION
U14993
VERSION
U14993.1
KEYWORDS
GI:550459
ORGANISM
Neisseria gonorrhoeae
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (sites)
Hagman,K.E., Pan,W., Spratt,B.G., Balthazar,J.T., Judd,R.C. and
Shafer,W.M.
Resistance of Neisseria gonorrhoeae to antimicrobial hydrophobic
agents is modulated by the mtrRDE efflux system
Microbiology 141 (Pt 3), 611-622 (1995)
JOURNAL
MEDLINE
95227358
PUBMED
7711899
REFERENCE
2 (bases 1 to 1500)
Pan,W. and Spratt,B.G.
Regulation of the permeability of the gonococcal cell envelope by
the mtr system
Mol. Microbiol. 11 (4), 769-775 (1994)
JOURNAL
MEDLINE
94254732
PUBMED
8196548
REFERENCE
3 (bases 1 to 1500)
Shafer,W.M.
Direct Submission
Submitted (21-SEP-1994) William M. Shafer, Microbiology/Immunology,
Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA
FEATURES
source
1..1500
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/strain="FA19"
/db_xref="taxon:485"
misc_feature
45..57
/function="protein binding target"
RBS
241..244
251..1489
/gene="mtrC"
251..1489
/gene="mtrC"
251..1489
/gene="mtrC"
note="mtrC"
/citation=[1]
/codon_start=1
/transl_table=1
/product="membrane fusion protein"

```


PUBMED 11823852
REFERENCE 2 (bases 1 to 215050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

COMMENT
FEATURES
source
1..215050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GM11000"
/db_xref="taxon:305"
1..11
/note="oriC or RS01885
/predicted by Framed"
/evidence=not_experimental
486..620
/gene="RS06178"
/note="synonyms: RSC0001, rpmH"
486..620
/gene="RS06178"
/function="structural elements; ribosome constituents;
ribosomal proteins - synthesis, modification"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Homology"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE 50S RIBOSOMAL PROTEIN L34"
/protein_id="CAD13528.1"
/db_xref="GI:17427009"
/db_xref="SWISS-PROT:Q8Y3H9"
/translation="MKRTYQPSVTRKRTHGFRVAMKTRGRAVLMARAKRRRLAI
"
gene
718..1113
/gene="RS01824"
/note="synonyms: RSC0002, rnpA"
718..1113
/gene="RS01824"
/EC_number="3.1.26.5"
/function="macromolecule metabolism; macromolecule
degradation; degradation of rna"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE RIBONUCLEASE P PROTEIN COMPONENT"
/protein_id="CAD13530.1"
/db_xref="GI:17427010"
/db_xref="SPTREMBL:Q8Y3H8"
/translation="MKTDERSVFALPRVRSRHFVLYVRANGHPQALGIVKKFA
RRAVERNLIKQCELEFLRLQALGGRVLLRLQTKPEREDVPTVAAFKRLCREBELSY
LFEIARPLPAPVPVPCPAALVPSDGAAP"
1110..1424
/gene="RS01825"
/note="synonym: RSC0003"
1110..1424
/gene="RS01825"
/function="miscellaneous; hypothetical/global homology"

gene
1431..3092
/gene="RS01826"
/note="synonym: RSC0004"
1431..3092
/gene="RS01826"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD13531.1"
/db_xref="GI:17427011"
/db_xref="SPTREMBL:Q8Y3H7"
/translation="MTRVLLFLRVVAVSPVGAQCRFLPTCSQDYARDAVLTGPA
IGSLAKRRCRCRHPFAQGVDPVPAAGDAALRSTDSASTETTAADAPARPSIHLP
RP"
1431..3092
/gene="RS01826"
/note="synonym: RSC0004"
1431..3092
/gene="RS01826"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD13532.1"
/db_xref="GI:17427012"
/db_xref="SPTREMBL:Q8Y3H6"
/translation="MDIKRTILWISLSVLLFDNMORANGHOSMFPPTORTTTA
APGCTPAGVPPKAAAPAAAGSQAAPRTGVSQTPASEKIVTTDVRATVDTRGAI
TKLELLTKDNDGPNWLPDRSLERTYLASGLIGDFPHNTVFLASAPRDGTG
EVSLLTADKDGAKLAKTYVFKRGSYVIDRFPVTDGAPINPTLMELARDGAVE
QSRYFTPTGPAVYTDHDKITFDIDSKSAVPAPDTSGWAMQHYFASAMIA
ASAKREYVDRIPTNFYRGMQALGVAGASVATARFAGPOEERMEIGTPE
LVMDYGLTIIAKPLLEKIKHLKGNMGSIVALTIVKLVFPLASYSYAMM
KDQDPRTAIRERKGDPOKMONBMTLITRYKENVLGGCLPIVIOIPVIALYVLL
SSVEMRGAPWLGWHDASDPPIILPTLWASVFTVTRLPDPDPVQAKMMFMPI
AFVMEFFPAGVLVWVNNCISIAQWISNMLTKKAAAK"
3222..4667
/gene="RS01827"
/note="synonyms: RSC0005, thdP"
3222..4667
/gene="RS01827"
/function="cell processes; protection responses;
detoxification"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE THIOPHENE AND FURAN OXIDATION PROTEIN"
/protein_id="CAD13533.1"
/db_xref="GI:17427013"
/db_xref="SWISS-PROT:Q8Y3H5"
/translation="MTASSHAMTSPVVSDAASAPRTVPIAATATAGRGIGVAVS
GPPVRAVMAQVGRLLPPRQATYLPFLDAGAIIDGIALMPAPHSYTEDVELOG
HGSPVWMLLSRCLRAGHGIGLVAPEGFTTRALNDLUDIAQAEVADLLLEAST
AARSARSIDGVESQTVHALVERVHLRLVEATLDPPEEIDPLEAADARQLADI
RAALDGVLAQARGLREGLHVLAVAGPVGKSLNLALAGELAIVPIACTTDX
VOOTIQEGIPINIVTAGRTEDEVERIGERTWAALARAVVHLIDPAAYRAG
LSAEDAIDARIEAHVPRGVPTIRVINKIDIAAGAVGVDAQPPVWASARGSGIE
LRAALLETINMGQGGEGYIARERLSALRSREHLLTAADADRAOSLDLFAEEL
RLAQEALNSTTGAFFSSDDLGLVIFSRFCIGK"
4757..5080
/gene="RS01828"

TITLE Stoneking, T., Nhan, M., Waterston, R. and Wilson, R. K.
Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

JOURNAL Nature 413 (6858), 852-856 (2001)

MEDLINE 21534948

PUBMED 11677609

REFERENCE 2 (bases 1 to 20382)

AUTHORS

CONSRITM

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

COMMENT

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and Genemark; <http://opal.biology.gatech.edu/Genemark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at Ecolyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladiada Salgado, Julio Collado-Vides and ReguonDB; http://knich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternative chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

location/Qualifiers

1..20382

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

195..729

/gene="accB"

/note="synonym: STM3379"

195..203

/gene="accB"

/note="putative -35 signal for accB; RegulonDB:STMLTH004589"

221..229

/gene="accB"

/note="putative -10 signal for accB; RegulonDB:STMLTH004589"

246..251

/note="putative RBS for accB; RegulonDB:STMS1H003389"

259..729

/gene="accB"

/EC_number="6.4.1.2"

/note="carrier of biotin; similar to E. coli acetylCoA carboxylase, BCCP subunit; carrier of biotin (AAC76287.1); Blastp hit to AAC76287.1 (156 aa), 92% identity in aa 1 - 156"

/codon_start=1

/transl_table=1

/product="acetyl CoA carboxylase, BCCP subunit"

/protein_id="AA122248.1"

/db_xref="GI:16421940"

/translation="MDIRIKIKLIELVEESISLEISSEGESVIRISFTTANAGFPVM QQAAAPMQQPALSNAAVAPATPAMEAPAAELISGHVRSFVPGTFRTPSPAKAF IEVQKVNVDLTCLIVEAMKMNQIEADKAGTVAIIVESQVPEFDEPLVIE"

gene 731..2089

/gene="accC"

/note="synonym: STM3380"

RBS 731..736

/gene="accC"

/note="putative RBS for accC; RegulonDB:STMS1H003390"

CDS 740..2089

/gene="accC"

/EC_number="6.3.4.14"

/note="similar to E. coli acetyl CoA carboxylase, biotin carboxylase subunit (AAC76288.1); Blastp hit to AAC76288.1 (449 aa), 96% identity in aa 1 - 449"

/codon_start=1

/transl_table=1

/product="acetyl CoA carboxylase"

/protein_id="AA122249.1"

/db_xref="GI:16421941"

/translation="MDKIVIANRGEIALRILRACKELGIKTVAVHSSADRLKHVLL ADETVICGPAPSVKSYLNIIPALISAETGAVAIHPGYGLSENANAEQVERSGFLF IGRKADTIRLMGDKVSAITAMKKGAVTPVSGDGLDDNNANRAHAKRIGVPVILIA SGGGGGRGMRVVSDAELAQSIEMTKAEAKASNDVMKEKLENFRHIEIOLADG OGNAIYLAERDCSMORRHOKVVEAPARIGITPLRRYIGRCARACACDYGAGTGE FLRENGEFTYEMNTRIQEHPTVMTITGDILKEQRIAGOPLSITQEBVYRGHA VECRINAEEDNTPLPSPGKITRHPAGGFVGRNESHAIYAGITVPYIDSMIGLICYG ENRDVAIRKMKALQELIDIGIKTINIDLOTRINNDHFHOGTNIHYLERKGLQEK"

2186..2440

/gene="yhdT"

/note="synonym: STM3381"

RBS 2186..2191

/gene="yhdT"

/note="putative RBS for yhdT; RegulonDB:STMS1H003391"

CDS 2198..2440

/gene="yhdT"

/note="similar to E. coli orf, hypothetical protein (AAC76289.1); Blastp hit to AAC76289.1 (80 aa), 82% identity in aa 1 - 80"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AA122250.1"

/db_xref="GI:16421942"

/translation="MDARFVOAHKEARMAWLTLCYAAWLVAAVYLPFGDSPGTLGPH WFMACLTPLVPLVILCWMAVVKFIYRDISLEDDAA"

2417..3881

/gene="panF"

/note="synonym: STM3382"

RBS 2417..2422

/gene="yhdT"

/note="putative RBS for panF; RegulonDB:STMS1H003392"

CDS 2430..3881

/gene="panF"

/note="similar to E. coli sodium/pantothenate symporter (AAC76290.1); Blastp hit to AAC76290.1 (485 aa), 93% identity in aa 3 - 485"

/codon_start=1

/transl_table=1

/product="SSS family sodium/pantothenate symporter"

/protein_id="AA122251.1"

/db_xref="GI:16421943"

/translation="WQLEVTLPVAVYLVVFGVSITYAMRRKTGTFINXYFLGSRSGC GIVAMTLTATVYISASSFIGPGAAKYRGVGLAMTQLPAWLSGLIGKKEFALIA RRYNAVTLNMLFARYOSRLVWLWLASLSLVARIGMTVOFIGARLLETNAGIPEYT GLIFGVISLVTYAFGGFRASVINDTLOGVMVGVITLVGVYIHAAGSGSQAVDTLH ALDPKIVTPQGADDITISPARMTSFVVLVCGVIGLPHRTAVRCISYSDSKAVHGIIIG TIVATITRMGMHAGLAGRAVYEDLTVYPDVITVLTWKKVLPRAAGIFLAAPMAAIS TIVAQILQSSATITIKDYLNLRPDQMNELRLRMSGATITLLGALLLLAKMPPEMT IWNLLAFGLAEAVFLMPLVLGLWBEANAGALSMTVGGVYVALLATENIYGLGH PIVPALILSLAFLIGNRFSSASQATVISTDR"

3893..4774

/gene="prnA"

/note="synonym: STM3383"

CDS 3893..4774

/gene="prnA"

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 11110)
Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I., Nelson, M.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J., Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J., Fraser, C.M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Ermeleva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.

Complete genome sequence of *Caulobacter crescentus*
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
11259647

2 (bases 1 to 11110)
Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I., Nelson, M.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J., Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermeleva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.

Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
SOURCE

Location/Qualifiers
1. .11110
/organism="Caulobacter crescentus CB15"
/mol_type="Genomic DNA"
/strain="CB15"
/db_xref="taxon:190650"
254. .721
/gene="CC0805"
254. .721
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="leucine-responsive regulatory protein, putative"
/protein_id="AAK22790.1"
/db_xref="GI:13422052"
/translation="MAISALDDTDRRLKVLQSDGRITNAELAKCNLSPATFDRVRRLRSGPTGYAALLDPKAVRLLIPVEVLKTTGDPFEEPGAAYRRAPELECHM VAGSPYLKARVADMEKAYRSLFDILVKMGVETRYVAVLEBKSTVOLPL"
complement (1149. .2552)
/gene="CC0806"
complement (1149. .2552)
/gene="CC0806"
/note="similar to GB:123839 PID:404773 PID:833173; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="efflux system protein"
/protein_id="AAK22791.1"
/db_xref="GI:13422053"
/translation="MLRNFILILVASTVAVGCTIAPQYRPVLPVAQTWSTPPEAATTVASADLDWRQVLDPRLOGVVDLALQONRLRVAVLIEKARQYGVORALPGVSGTAVGEORGRTPASTPTGQPEPTDFVFNATIGFTAYELDFGRVSLDKDALQSLATEQTSSTSVSLIAEIANAMLTIAADOERLAKTTLATEDESLRVRORVDGASGLDPLRNMTETLAETRADAVAVYTAQVADRNLTLIAGDLPVLDLPPRLDADSVLADLPA GLPSDVALRPPDVAAEROLQANANITGAARPAAPRSLTGSAGSASTDIDGFKSGTSTSTFTPOLSTPTFAGSANYANLNTISGADDIIVATYEKTVAGTARFREVSPALSRATVADRLLTQERLTAAADSTRLSQQRDGLDLSALTLLDAQSLVTAQGLISARLAQATNLVLYKTLGGCAPT"
complement (2554. .5709)
/gene="CC0807"
complement (2554. .5709)
/gene="CC0807"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="AcetB/AcrD/AcrF family protein"
/protein_id="AAK22792.1"
/db_xref="GI:13422054"
/translation="WLSRFITDRPIFAWVIAIVIMLAGALAIRLPINQVPEIALPOVSISANYPASAKTVDSTVQVIEQMKGLDGLDWSSTSDSGSATVTLTFKAGTDIDIAQVQVOKLOTATALLPOEVQOQGLTVAKSARPMVMVIGLSEDPKTTGADLADYMA

gene
CDS

SNLQDPLSRVDVGDIQLFGSQYAMRIWLDPOKLAESLTPADVAAAIQAQNAQVNSG QIGCTFNLPPTGATNATITQASRLQTEBOFOIIVKNTSGCATLDDVAVELGAEYS ASIAKNGXQAAQMAIKLAKAGNALPTAAVNAVRMOLKNGPANKYVLPYSTFEV KLSIEVYITVIAITLIPVIMFLPIQNRATLIPVAVVILGTEGVAAAGSINS TLMPFGVLAIGLVDAIVYENVRVNSBESLSKEATRKSMNETGLITGLIAYLV AAFVFPMAFFGSGQVITYOFSTITVSAMLSVVVALVLPALATMLKPKVKGHHE KTFPFPMFNKSPFDMSSRYQSVRKILGKRGRWMAIYALIVMGLFVLPSPAFLPE EDOGTMTFTVLQLPAGATTEKTLVLDQVRDHPFLVGEKAVOSVFTVSGFSFAGQNA GLAFVRLKDEPHERKANLKAQAVAGRAMGAFSOFQDAMVAVIYPVAVQELGSSGDFP QLODVGCVGHEALNARNMMLGMAQDPSLVGVRPNQODPTPOLKTEVDQAKGAGLT TTDINDLSAAMGSGVYNDFTDRCGVKKYMGADAAPFMTPEPDLNRMVYRNQGVY AWFAPFATSWTISPPRLERYNGISSVNIQCAPRPGSSGALIMAKILAKLPQVY EFWGASQELAEANOAPALYALISLVFILLALYBSWSIPLAVINXITPLUGVIGALLA TFRGGLSNDIYFPGVGLITWGLAKKAILIVERAKDLYEKQMGILKATLEAVRLRLEP IMTSLAFVYGVPLAISNGAGSQAQAIQGVIGCMISATLLAIFVPLFVVEKIK FKXKNGHDAQGVGSTERH"
complement (5722. .6888)
/gene="CC0808"
complement (5722. .6888)
/gene="CC0808"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="HlyD family secretion protein"
/protein_id="AAK22793.1"
/db_xref="GI:13422055"
/translation="MHPQREVAATTVQGLAIALITLACGQPGSGMGMGWGPPEVXY IYVQSSVGLSTELAGRTSYLYSEVRPOVSQVYQRLPEEGLVAVRAGQPLVQIDPAT YQAAVYSAALALAOQAOATPAKULKADRIALVETGANSQODDDADAAALQTLAAVY VQKAAALDSARININVARVTAPISGRIKSSVTALGAVTANQALATVQDLSYVLD TQTSABELIKLQAPFASGKVSQAVTLLLEBGSYPIPIGRLEFSDITVDPGGAAG LRAVFDNPKGVLLPGMYRAVLDSQVAPSGIILPQAVNDDPGKATVMVYGAKKGPP RPVTLQGVTDKMLVTSGLNADGKVIYEGMKVRPGAPVAVPAGAAPAAQAQR"
7182. .7670
/gene="CC0809"
7182. .7670
/gene="CC0809"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Tetr family"
/protein_id="AAK22794.1"
/db_xref="GI:13422056"
/translation="MAEIAKAGLSVQVIRYFENKEATITAAVADOLAMERDKFAEL ERPGTIRDSIDHLEPAIDKCFDERRALALVVAEARNPKVAIPVAIRADAOERL ARAAMBERDKRPESSAEFARCEVGLTFGLLRGVNHRDLDRDLTDLVLRSTVRL LS"
complement (7667. .8449)
/gene="CC0810"
complement (7667. .8449)
/gene="CC0810"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22795.1"
/db_xref="GI:13422057"
/translation="MRTITYTDAEPDNTQALASSRYLAADDOPLGDSMRGVRFILE FAKADELARWGVRSITIVFGSARVERDGEPEERARVTAERLGRITVISEGAGHQNG DKVRDNVIATGGGPGMLEANRGAFFEGASIGFNTILPHEOVNPPVCTEPLTFRFY FAKRKMLARANALVFPFGSGTFDELFEILRLRQDKAPPIPIVLFDEAYRSIIN FEALVEHGMTAASDLBELVHPADDAEAVMDKLVDCGLPLPGV"
8564. .10156
/gene="CC0811"
8564. .10156
/gene="CC0811"
/note="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="mecalio-beta-lactamase family protein"
/protein_id="AAK22796.1"
/db_xref="GI:13422058"
/translation="MTSITLPHGAGCVTGCARLVTPRTVVLIDCGMFGQSKTKAL

TITLE Kitajima, J.P.
Comparison of the genomes of two Xanthomonas pathogens with differing host specificities

JOURNAL Nature 417 (6687), 459-463 (2002)

MEDLINE 22022145

PUBMED 12024217

REFERENCES 2 (bases 1 to 13945)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Camarvan, F., Cardozo, J., Chamerling, F., Clapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L., Cusimbo-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Taita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

FEATURES Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

TITLE JOURNAL

SOURCE 1. 13945
/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
/mol_type="genomic DNA"
/strain="ATCC:33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
151. 1854
/gene="XCC2674"
151. 1854
/gene="XCC2674"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="extracellular serine protease"
/protein_id="AAM41946.1"
/db_xref="GI:21113852"
/translation="MLAQTPTATSRAPQRPPESDPLFRYQHLNDGQEVATDTPRA GTDLNVPGLTLGRGQVTAVVDDALELHPDLVANVVGSGSKNPDGSDPTPRN GASHGTWVGIMAAVGNGLGARGVAPARLKNILSSSESPDPTDNLRAAGSDVY OSRDVDNNSFGSDLTYPPTISPAEBSIDLRMRBRANGKGLYVOAGNTPDSFTY LDDGQNVVERCPVARTIGVTCSTPATDPLISNOLITATGAVNARGLRSSVSSGAL WYTFGGEFGIQRKRYFGDRPPALFDPATVITDITGCAVGNRIAGCPINLASSS SPIDASCNYSAAWNGTSAAPTVAGVALIVQANPSLTARDLKITLATASQIDPAQO VIRYQGVIEPQWITNAGHAFSNWYGLVDAEAVYRAYGQPLPORDLGMKAR AGTGTIGPNPAATLRLRIGDTLKIDTVQSMOTHTKPSNLRVILISPSGTSRYVLT PQALDITLQAGFEIPLSTSNAPLDEWVAGTWLLEVTDMGSDVPAGLTFEKLKING H"

gene 1856. 2446
/gene="XCC2675"
1856. 2446
/gene="XCC2675"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM41947.1"
/db_xref="GI:21113853"
/translation="MNALICTLLICALSWPASAOPTDRLAOPAAAGTTCIDPN FRLVFNARVLAINTPPPRADVPANRSTTAAGTTLCHTIGYVSSGNSPPTTGY GVEEPLMVAIINGNGQPVVASPRLQVYADADAIVKLAQTNGAVLAASQDGTGYI GYRSVPEALALPTVRGGVDVVLQVIAVPRLR"

CDS 3031. 3942

CDS 3031. 3942
/gene="XCC2676"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="transcriptional regulator"
/protein_id="AAM41948.1"
/db_xref="GI:21113854"
/translation="MASSIMHLVDQNRHAPTOGTSTPAVPGIRIMCSPATYSPV YTPPLIAVQGHQKQALGTYVHHPAMVYVASVDLPIISGVIBASPOYPLCCLD DVAVLSLALRYPALAEQAGDTPASGLINRSTPELDAALRLARLQDPRDIDALP LVIRELLRYMAEFGNAIVEQMAIADSRILQISRAIMLAHYAERFSIEQIDLSAM GSSTPHFPAVMSPTPEFSHRLRVEARLIMVAEALSAADAGFRVYESASQFSRD YARLIGPPARDQQLRASRODAIEVA"
complement (4154. 4591)
/gene="XCC2677"
complement (4154. 4591)
/gene="XCC2677"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAM41949.1"
/db_xref="GI:21113855"
/translation="MAAGRGADENHAPIKVTXGASASAGPASFTGTVRIDAPPOS DAPARVGATVTFEPGAKRTKWHTHPLQOTLIVTGRAGRVQEMKRPADIEIRGDIWLP PGVAMHGNATVAMSHIALAEAVDSPTWLEQVSDAQGN"
4794. 5708
/gene="XCC2678"
4794. 5708
/gene="XCC2678"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="transcriptional regulator lysr family"
/protein_id="AAM41950.1"
/db_xref="GI:21113856"
/translation="MWSASWARENLDLQVFTVAREGSTTRAAAGLQVSQALSHTV RALETRGIRLITRTSRVSMTEAGMKLPPTVAPRLEIDQESALTAEPDKPAGTIR ITAGHAAADAYLWLPALSKVLPDYDEITVDVGLADIVAEKYDGLRGDQVAKDM IAVPISPRQMAVVAASPFVHTTIPAVPRLAONHNGIGRLPFGHGLPMDPEKQDR EOKRVNDGWTQFNSSAMLRBAALAGLAPLPETWLDHIAAGLRVLDMDKCPDFG YIAYPERRQSSAAKVTVDALRHDLVR"
complement (6036. 7019)
/gene="XCC2679"
complement (6036. 7019)
/gene="XCC2679"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="oxidoreductase"
/protein_id="AAM41951.1"
/db_xref="GI:21113857"
/translation="WHTRTLGRSGLRVSAIGLGMGLSYTGATERTKQVALLHAAY EOGVTLFTDAEAYPIANBELGEBALVGAQVAVAKFGRGHADAGLDSRAPLR AVAEASLRLRKRERIDLFYQHRVDPAVPIEIVAGTVADLTAEGVGHFGISEASARIT RRAHAAQVPAVAVQSEYGLMWRPEASVLPLEELGIGFVFPSPGKALFGALQOPDQ FQDSDPNSVPRFAEABRRANQALVQGLAIDDKGATPPQVLAWLMLTSKPMVPIP GTTKRRLDENIGALVSSDELARIQALDAVEIYGERXSPERQVLVK"

gene complement (7291. 7980)
/gene="XCC2680"
complement (7291. 7980)
/gene="XCC2680"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="transcriptional regulator"


```

Db      1919  GTTGAACCGCGCGGTATCACTGCTTATATCAAAAGTAGCTGCCCATATCATGAGCGC 1978
               |||||
Qy      191  IleglyInserLysValSerGluGlyThrLeuLeuAnaIaGlyAspThrThrValLeu 210
               |||||
Db      1979  ATCGTAAATCCGGCGGTACCGAAGCGCGGTGCGCAAAACGGGCATCTACCGCGCTG 2038
               |||||
Qy      211  AAlaThrIleArgGlnThrAsnProMetTyValAsnValThrGlnSerAlaSerGluVal 230
               |||||
Db      2039  GCTACCGCTTACAGACGTACCTATCTATGTTGACGTTACCATCGACAGACGATTTCC 2098
               |||||
Qy      231  MetLysLeuArgGlnIleAlaGluGlyLysLeuLeuAlaAlaAspGlyValIleAla 250
               |||||
Db      2099  CTGGCTGTGAACAGGAACCTGGCGAACCGTAACTGAAGCAAGAAACGGTAAAGCGGAAA 2158
               |||||
Qy      251  ValGlyIleLysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPheAla 270
               |||||
Db      2159  GTTAACTGCTTACCATGACGCTCTGAATATATCGCAAGAGGAGTACGCTGAATCTTCC 2218
               |||||
Qy      271  AspProValValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
               |||||
Db      2219  GACGTCACCGTTGACAGACCGGTTCTATCAAGCTGGCGGCTATCTCCGACCCG 2278
               |||||
Qy      291  GlnAsnIleLeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaValAsp 310
               |||||
Db      2279  GACCACACTGCTGCGCAGGAATGTTGTTGCGCACGCTGGAAGAAAGTATTAAACCG 2338
               |||||
Qy      311  AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaAsp-----ThrVal 328
               |||||
Db      2339  GATCGCGCTGCTGCTTCCGCAACAGGCGTTTACCGCTACCCCGCGCGGATGCCAGCGTG 2338
               |||||
Qy      329  MetIleValaAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGln 348
               |||||
Db      2399  ATGGTGTGCGCGGAAGCGCAAAAGTTGAAGTGGCCAGCGTCAACCGGACTCAGCGTATC 2458
               |||||
Qy      349  GlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValValGluGly 368
               |||||
Db      2459  GGCACAAAGGTGTAGTACCGAAGGCTGTGAATCAGCGCGTCCGCGATGCTGCTGCG 2518
               |||||
Qy      369  IleSerIleAlaGlyIleThrGlyAlaLysLysValThrProLysGluTrpAlaSerSer 388
               |||||
Db      2519  CTG-----CAAAAGTCAAAACGAGCGTACAGGTAAAGCG 2554
               |||||
Qy      389  GluAsnGlnAlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAla 408
               |||||
Db      2555  CAGGAAGTACGTTCAAGCAACCAACAG-----CAAACCGCGGCAACGCAAGCTCTCAG 2608
               |||||
Qy      409  SerGlu 410
               |||||
Db      2609  TCTGAA 2614
               |||||

RESULT 14
ECONCRAB      4879 bp  DNA      linear  BCT 11-JUN-1993
LOCUS      E. coli acriflavine resistance proteins (acrA and acrB) genes,
DEFINITION      complete cds.
ACCESSION      M94248
VERSION      M94248.1  GI:290404
KEYWORDS      acrA gene; acrB gene; acriflavine resistance protein; cytoplasmic
SOURCE      membrane protein.
ORGANISM      Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE      1 (bases 1 to 4879)
AUTHORS      Xu, J. and Bertrand, K.P.
TITLES      Nucleotide sequence of the acrAB operon from Escherichia coli
JOURNAL      Unpublished (1992)
COMMENT      Original source text: Escherichia coli (strain K-12) (library:
              Clarke-Carbon) DNA
FEATURES
    source
        1..4879
        /organism="Escherichia coli"

```

ORIGIN	BASE COUNT	1157 a	1215 c	1343 g	1164 t
Alignment Scores:					
Pred. No.:	2,21e-36	Length:	4879		
Score:	762.50	Matches:	168		
Percent Similarity:	61.31%	Conservative:	76		
Best Local Similarity:	42.21%	Mismatches:	129		
Query Match:	37.77%	Indels:	25		
DB:	1	Gaps:	6		
US-09-889-756A-2 (1-412) x ECOACRAB (1-4879)					
OY	11	AlaAlaAlaLeuAlaAlaAlaValaLeuValaLeuSerSerCySGlyLySGLyglYasp	30		
DB	359	GGCGTCGTTCTGATGCTCTCCAGCGACGCTTACGCCCTTACAGATG---GACGACAAACAG	415		
OY	31	AlaAlaGInGlyglYglInProAlaGlyArgGluAlaProAlaProValaValglYVala	50		
DB	416	GCCCAACAAGGGGGCCAG-----CAGATCCCGCGTGGCGTAGTA	457		
OY	51	ThrValHisProGInThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu	70		
DB	458	ACAGCAAAACGTGAACCTTCGACATCAACACCGACTTCCGGGTCGACACAGTGCCTAC	517		
OY	71	ArgThrAlaAspValaArgAlaGlnValaGlyglYleileGlnLysArgLeuPheGInGlu	90		
DB	518	CGATCCGACGAAGTTCCTCTTAAGTTAGCGGATTAATTCGAAGCGTAATTCAAAGAA	577		
OY	91	GlySerTyValaArgAlaGlyglInProLeuTyrgInleAspSerSerThrTyrgLuAla	110		
DB	578	GGTACGACATCATGACGACAGGTGTCTCTCTCATGATGATGATCTCGACATCATCAGCG	637		
OY	111	AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaAlaValaAsp	130		
DB	638	ACATACGACAGTGCACAAAGTGATGTGGCGAAAGCCACAGGCTGACGCCAATATGCGGAA	697		
OY	131	AlaAspLeuAlaArgTyTyLysProLeuValAlaAlaGluAlaValaSerArgGInGluTy	150		
DB	698	TTGACGGTGAACCTGTTATTCAGAAATGCTCGGTACTCATGATCATCATGTAAGCAAGATAC	757		
OY	151	AspAlaAlaValaThrAlaLysArgSerAlaGluAlaGlyValaLysAlaAlaGlnAlaAla	170		
DB	758	GATCAGCGCTCGCGATGCCGACAAAGCGAGTGCGGGTAACTCGCGCAAAACCTCC	817		
OY	171	IleLysSerAlaGlyLleAsnLeuAsnLysArgLleThrAlaProIleSerLysPhe	190		
DB	818	GTTGAACCTGCCGATCATCTGGCTCTTACACCAAAATCACTCTCCGATTAGCGGTGCC	877		
OY	191	IleGlyGInSerLysValaSerGluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu	210		
DB	878	ATTGGTAAAGTCGAACGTACGGAAGCGGATTTGTAACGAAGCGTCAGCGCATCGCGTG	937		
OY	211	AlaThrIleArgGInThrAsnProMetTyTyValaAsnValaThrGlnSerAlaSerGluVal	230		
DB	938	GCAACCGTGCACGAACCTTGATCCGATCTACAGTTGATGACCCACAGTCCAGCAACACATTC	997		
OY	231	MethLysLeuArgArgGlnIleAlaGluGlyLysLsLeuAlaAlaAspGlyValIleAla	250		
DB	998	CTGCCCCCTGAACAGAGAACTGGCGAATGGCAGTGAACAAGAAAGCGCAAAACCCAA	1055		
OY	251	ValGlyIleLysPheAspAspGlyThrValTyTyProGluTySGLYArgLeuPheAla	270		
DB	1058	GGTGCATCATGATCAACAGTGAAGCGGATTAAGTTCCCGCAGGAAGGTGAAGATTCTCT	1111		
OY	271	AspProValaValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValaProAsnAsp	290		
DB	1118	GACGTTACCGTTGATCAGACCACTGGTCTATACACCTTAACGCGTATCTTCCCGAAACCG	1177		
OY	291	GlnAsnIleLeuMetProGlyLeuTyTyValaArgValaLeuMetAspGlnValaValaAsp	310		
DB	1178	GATCACTCTCTGCTCGGGATATGTTCTGCGCGACGACGTCTTGAAGAAAGCGTTATATCA	1233		

OY		311	AenlaaprevlvalProginglnAlalValThrprgGlyAlalyasp-----ThVal 328 ::: 	
Dd		1238	AACCTATTATTTAGTCCCGCACAGGGGTAAACCCTAACGCCGCCTGGCAGTCACCGTA 1297 ::: 	
OY		329	MettilevAlasnalaglnglgylMetcgluProlargluVallthrValAlaglIngln 348 ::::: : 	
Dd		1298	CTGTACTTGtggcgccgatgcACAAGTAGGAACCCGTCCCATCTTGCACAAGCCAGCTATT 1357 ::: 	
OY		349	GlyIhrantppillevAltHrserGlulelyspsglyAspySvlvalvalIgUngLy 368 ::::: 	
Dd		1358	GGCAATAAgttccttgtagAcgaAgcttGtaaacacaggcaatgccctaTaataaaGracc 1417 ::::: 	
OY		369	IleSerllealaGlyIlernghyAlalySvlvalThrpro----- 382 ::: : 	
Dd		1418	CTG-----CAGAAGTGCGTCTGTGTGTCACGTAAMAGCA 1453 ::: 	
OY		383	LysclutRpalaseSerGluErghnAlalAlaalProginserGlyValGIn 400 ::: 	
Dd		1454	CAMAAGTTAccctGTGaTAAAtAACCaCAAGCCGCA-----AACGGTGTCTAG 1501 ::: 	
RESULT 15				
E27981		E27981	4879 bp DNA linear PAT 18-JUN-2001	
LOCUS				
DEFINITION		Organic solvent-tolerant Escherichia coli and process for producing the same.		
ACCESSION		E27981 E27981.1 GI:13018304		
XREF		JF_1999221080-A/1.		
KEYWORDS		unidentified		
SOURCE		unidentified		
ORGANISM		unclassified. Rikizo,A., Toru,H. and Toshiaki,K. Patent : JP 1999221080-A 1 17-AUG-1999;		
REFERENCE				
AUTHORS		Rikizo,A., Toru,H. and Toshiaki,K.		
TITLE		Organic solvent-tolerant Escherichia coli and process for producing the same		
JOURNAL		Patent : JP 1999221080-A 1 17-AUG-1999;		
COMMENT		MEIJ I SEIKA KAISHA LTD OS Unidentified PN JF_1999221080-A/1 PD 17-AUG-1999 PF 09-FEB-1998 JP 1998027537 PR RIKIZO AONO,TORU HAMAYA,TOSHIAKI KONO PT C12N15/09,C12M1/21/(C12N15/09,C12R1:19),(C12N1/21,C12R1:19),PC C12N15/00, PC C12N15/00,(C12R1:19) PC Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers FT CDS 329..1522 FT sig peptide 329..400 FT mat peptide 329..1522 FT CDS 1530..4694. FT CDS Location/Qualifiers 1..4879 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"		
BASE COUNT		1157 a 1215 c 1343 g 1164 t		
ORIGIN				
ALIGNMENT SCORES:				
Pred. No. :		2,21e-36	Length:	4879
Score:		762.50	Matches:	168
Percent Similarity:		61.31%	Conservative:	76
Best Local Similarity:		42.21%	Mismatches:	129
Query Match:		37.77%	Indels:	25
		6 Gaps:		6
DB: US-09-889-756A-2 (1-412) X E27981 (1-4879)				
QY 11 AlalAlalAleuuAlalAlalAlalAlalAlaleuValLeuSerSerCySGlyLyrGlyGlyASP 30				

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE
1 (bases 1 to 12524)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,B.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,U., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
NATURE 409 (6819), 529-533 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
11206551

REFERENCE
AUTHORS
2 (bases 1 to 12524)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,B.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,U., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 12524
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
complement (110. .619)
/gene="y1aC"
/note="synonym: Z0570"
complement (110. .619)
/gene="y1aC"
/function="orf; Unknown function"
/note="Residues 1 to 169 of 169 are 100.00 pct identical to residues 1 to 169 of 169 from Escherichia coli K-12 Strain MG1655; B0458"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG54807.1"
/db_xref="GI:12513325"
/translation="MPEPKIQRPRQMTETIQRLTETISLNTREKDKKPPRSISFIRKPGIPIGMYVAFATLAVMLQSEITLGSVWLVLVLLGFFFDVPRYREDIDVLDPRVCYNGEWMYTRFVPALVEALINSPRVADVHKEQLKMIVRKGLSEFYDIFTLRAESTS"
complement (696. .1247)
/gene="y1aD"
/note="synonym: Z0571"
complement (696. .1247)
/gene="y1aD"
/function="putative enzyme; Not classified"
/note="Residues 1 to 183 of 183 are 99.45 pct identical to residues 1 to 183 of 183 from Escherichia coli K-12 Strain MG1655; B0459"
/codon_start=1
/transl_table=1
/product="putative transferase"
/protein_id="AAG54808.1"
/db_xref="GI:12513326"
/translation="MSTREKMTAGELYRSADETLSRDRIARQLIHRVHSLAEHTLRQILADLFGQVTEATIEPTFRCDDYINIFLGNSFANDCVMLDVCPIRIDNCMLALGVHLYTATHPIDPVARNSGAELGKRVITIGNVMVIGRAVINPGVTIGNVVAVASGAVVTKDVPDNNVVGNPARIITKL"
complement (1419. .1637)
/gene="y1aH"
/note="synonym: Z0573"
complement (1419. .1637)
/gene="y1aH"
/function="regulator; Macromolecule synthesis,

modification: Proteins - translation and modification"
/note="Residues 1 to 72 of 72 are 100.00 pct identical to residues 1 to 72 of 72 from Escherichia coli K-12 Strain MG1655; B0460"
/codon_start=1
/transl_table=1
/product="haemolysin expression modulating protein"
/protein_id="AAG54809.1"
/db_xref="GI:12513327"
/translation="MSRKPLTKTDYLMRLRCQTIDTLERVIENKRYELSDNELAVFYSAADHRLAELTNMKLYDKIPSSVWKFR"
complement (1663. .2037)
/gene="y1aJ"
/note="synonym: Z0574"
complement (1663. .2037)
/gene="y1aJ"
/function="orf; Unknown function"
/note="Residues 1 to 124 of 124 are 100.00 pct identical to residues 1 to 124 of 124 from Escherichia coli K-12 Strain MG1655; B0461"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG54810.1"
/db_xref="GI:12513328"
/translation="MDEYSPKRDIAQLKFCETLYHDCILANESNHNQVNDPTSAINQINLEIHLATFALNLYKIKYNEDEKLLIQIDIEYLDPTFMLSISYINNOIQKMKSGNRLPRCPVNAKKNPAGLSC"
complement (2583. .5732)
/gene="y1aCB"
/note="synonym: Z0576"
complement (2583. .5732)
/gene="y1aCB"
/function="transport; Protection responses: Drug/analog sensitivity"
/note="Residues 1 to 1049 of 1049 are 99.80 pct identical to residues 1 to 1049 of 1049 from Escherichia coli K-12 Strain MG1655; B0462"
/codon_start=1
/transl_table=1
/product="acridine efflux pump"
/protein_id="AAG54811.1"
/db_xref="GI:12513329"
/translation="MPNFRIDRPIFAMVIAIITIMLAGGLAILKPVAYPTIPAYTISASYPGAKTYODPTOVITBONNAGINIMWMSNSSTGVOITLTPESTDDIAQOVQNKIQAMPILPQEVQOQVSIEKSSSFLWVGIVINDGIMTQEDISDYAA NMKDALSISGVDVOLFGSOYAMRIWMNPNELNKFQLPVDVITAIKQNAQVAAQ LGGTPVVKGOQLNAAIIAQTRLTSTEEFGKILKVNQDGSRVILRVAKIELEGEND IIAEFGOPASGGIGIKLATGANALDPAARAIKALAMPEPFGKIVVYDTPFVK ISIHVVKTILVEAILIVPLVMVLFIONPRAITLPTIAVPVILGTAVALAAGRSINT LTMFGVLAIGLVDALIVVERVERVMAEBGLPREAKTRKSGQIOGALVGIAMTUS AVFVPVAFPGSGTALYRQPSITIVSAMLSVVALITPALCATMLKPLAKODHSGG KGGFGWFMNMEFKSTHAYTDSVGLITRGRVLYLIIIVMAALFARLPSSFLPD EDGQVFMTVOLPAGATOERTQGLNEVHYITKKNVSVFVANGGFAFGCGNT GIAFVLIKMADEPGENEYVEAITMRAFAFSQIKDAMFAFVLPVIVELGTAGDF ELIDQGLHEKTLQARNOLLAAAKHPMLTSVRNGLEDPPQKIIDDOEKAOLG VSINDINTLTGAAGGSYYNDPTDRGRYKVVYMSAKTRMLPDDIGDNYVRADQM VPSAFSSSRMEGSRRLRYNGLPSEWITLGOAPKSGTEAELEMLQSLAKPLTGG YMTGMSYQERLSGNQPSLYAISLVILCLAALVESIPSVMLVPLVIGLIL AATFGLITNDVYFOVGLITTLTIGASAKNALIIVFADLMDKEGKILTEITLDVVRRL RPLMTLSLAFILGVLPIVSTGASGAQNAVGVGVGVTATVALIIFVVPFVVVR RRSRKNEDEIHNNHTVDHH"
complement (5755. .6948)
/gene="y1aCra"
/note="synonym: Z0578"
complement (5755. .6948)
/gene="y1aCra"
/function="transport; Protection responses: Drug/analog sensitivity"
/note="Residues 1 to 397 of 397 are 100.00 pct identical to residues 1 to 397 of 397 from Escherichia coli K-12 Strain MG1655; B0463"
/codon_start=1

ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 13446)
AUTHORS Blatter, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Coliado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Ma, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)

TITLE JOURNAL
MEDLINE 97426617
PubMed 9278503

REFERENCE 2 (bases 1 to 13446)
AUTHORS Blatter, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

REFERENCE 3 (bases 1 to 13446)
AUTHORS Blatter, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

REFERENCE 4 (bases 1 to 13446)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHER). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'p' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES

source Location/Qualifiers
1. 13446.
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
1. 22
/note="factor Sigma70; predicted +1 start at 475839"
86. 439
/gene="ybaA"
/note="synonym: b0456"
86. 439
/gene="ybaA"
/function="orf; Unknown"

gene
promoter
CDS

/note="orf; 100 pct identical to YBA_ECOLI SW: P09161"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73558.1"
/db_xref="GI:1786661"
/translation="MKVDFVVAVPADKDAVREMAKAPLFEKFGALRIYECMAS
DVPDGKVDPRMAVKAENEVVFWSWIEYPSKEVRDANQKMSDPMKKEGSEMPD
GKMIVYGFESIIDE"
complement (481. 2037)
/gene="y1aB"
/note="synonym: b0457"
complement (481. 2037)
/gene="y1aB"
/function="orf; Unknown"
/note="f518; This 518 aa ORF is 33 pct identical (6 gaps)
to 513 residues of an approx. 536 aa protein Y0CC_ECOLI
SW: P32701"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73559.1"
/db_xref="GI:1786662"
/translation="MLVRIHLVGLISGLILSLPVLGSLVLAHQVETSPTEILD
TVSSRAIRANKVATQKALQELERQACSAHIMEMRVSRYIQEVAYIDN
NVPQCSLEHSPDTPPEBPQKISQDQRYVSLGIRVWAGTAYVIMDP
ASFDIVIPYSWQIDAAIIGNANVVTSDSDTAQGITLQKTPGSHIENGIIYDI
LPLPENWISITWASTKMLQKGMHROVFTLPGIVGLAAMPVLRIRIOSPHR
LQALINRDIQVHQPVSILANGKIYGAELAMPPODGSWLSDSITPLAQOGLSE
PLTLIRSVFEDMGMLRQHPQOHSINSESVLTSEKIPLQLRDMINHYOVNROI
ALBLTERFADPPTISAPILSRVAGHEIYLDPGTGSLSYQDIDVILKIDKSF
VDALKEVKNVPHIEMAKTLKLVNAGEIETSKQEWLROHGVHYGGWLSKALPRE
DFRMAEOHL"
complement (2052. 2080)
/note="factor Sigma70; predicted +1 start at 477855"
complement (2195. 2704)
/gene="y1aC"
/note="synonym: b0458"
complement (2195. 2704)
/gene="y1aC"
/function="orf; Unknown"
/note="f169; This 169 aa ORF is 32 pct identical (0 gaps)
to 53 residues of an approx. 592 aa protein ASN1_PEA SW:
P19251"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73560.1"
/db_xref="GI:1786663"
/translation="MPGKLOIPRROTMTIORLTTETTESLNTREKRNKPRFSISF
IRKHPGLPIGVAVAPATLAVMIQSETLSGSWLVLVLTILNGPFPDVPYPRRYED
IDVLDPRVCYNGSWYNTFRVPALVBAIILNSPRVADVHKEQLQKMLVRKELSPYDIF
TLARASTS"
complement (2708. 2735)
/note="factor Sigma70; predicted +1 start at 478511"
complement (2722. 2750)
/note="factor Sigma70; predicted +1 start at 478525"
complement (2781. 3332)
/gene="y1aD"
/note="synonym: b0459"
complement (2781. 3332)
/gene="y1aD"
/function="putative enzyme; Not classified"
/note="f183; This 183 aa ORF is 65 pct identical (2 gaps)
to 183 residues of an approx. 192 aa protein YV1_BACSU
SW: P37515"
/codon_start=1
/transl_table=11
/product="putative transferase"
/protein_id="AAC73561.1"
/db_xref="GI:1786664"
/translation="MSTEKEXMIAGELVRSADETLSRDRLRARQLIHRYNHSLAEHT
LRQQLIADFGQVTEAYIEPTFCRDYGVNIPFGNPFANFDVCMADVCPRIIGDNML

promoter
gene
CDS

ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 139818)
Roberts, D., Allen, E., Araujo, R., Aparicio, A., Chung, E., Davis, K.,
Duncan, M., Federspiel, N., Hyman, R., Kaiman, S., Komp, C., Kuri, O.,
Lew, H., Lin, D., Namath, A., Oetner, P., Schramm, S. and Davis, R.W.
Sequence of minutes 4-25 of Escherichia coli
Unpublished

TITLE JOURNAL
2 (bases 1 to 139818)
Roberts, D., Allen, E., Araujo, R., Aparicio, A., Chung, E., Davis, K.,
Duncan, M., Federspiel, N., Hyman, R., Kaiman, S., Komp, C., Kuri, O.,
Lew, H., Lin, D., Namath, A., Oetner, P., Schramm, S. and Davis, R.W.
Direct Submission
Submitted (19-DEC-1996) Department of Biochemistry, Stanford
University, Stanford, CA 94304, USA

FEATURES
source location/Qualifiers
1. .139818
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
/map="minutes 9 to 11"
1. .684
/note="overlaps the sequence in GenBank Accession Number
U73857, which covers the 6 to 8 minute region"
<1. .360
/note="similar to S. typhimurium branched chain amino acid
transport system II carrier"
/codon_start=1
/transl_table=1
/protein_id="AAB40157.1"
/db_xref="GI:1773085"
/translation="LGGFMSVNVSLGLSOLISVPLTATVPCIALVVLSPFRSW
HNSRVLAPPMFTSLFGILDGIAKSAFSDILPMAQRPLAELGALMLPTVWVVL
AIIYDRAAGROVSSAH"
436. .1809
/note="similar to S. typhimurium Prox"
/codon_start=1
/transl_table=1
/protein_id="AAB40158.1"
/db_xref="GI:1773086"
/translation="WESKTKLKRGSLSTRHIFMAAGSAGITGLFYGSADAIKAGPSV
LADVITIGIAVYIMRALGEMSVNPPASSFSRYAENLPLAGYITGMTCYCEILIV
LIIADYVFGIYMGVPTPHMIVLVSLLICAVNLMSYKVGLEFNFSPKVLAT
IIMIVAGEGIIIMGIGNGOQTGTHNLMSNGFSPNGMLGMMSLQMWMPAAGYIEI
GITAGEAKDEKSIIPRAINSVPBRIIVFYVGLTFLVMSIIPMNQVGTAGSPYLITQH
MGITFAASITINPVUTASLSAINSDVGVGRMLHGAEGSAPKIRSKTSRGIPIWT
LIVWTTALFAVYLYNITMPENVFLVLAISLATFATVWVWIMILSQAIFRRRLPEEVK
ALKFKVGVATTIGLIFLFTLIGLIGYHPTDRISLYGVFAWIVLLIGMFKRRHD
ROLAENG"
1680. .4024
/note="previously sequenced by another group and can be
found in GenBank Accession Number X59839"
1965. .3782
/gene="malZ"
1965. .3782
/gene="malZ"
/codon_start=1
/transl_table=1
/product="maltodextrin glucosidase"
/protein_id="AAB40159.1"
/db_xref="GI:1773087"
/translation="MMNLAMHLVPPRPVKOSKOLLITLMTGDEPPORTMLTEHDN
EMKSVPMHQRSGQPGVGTAWRAIDLSQSPRRRSFKLIMHROBMFLTPQGSMP
PALBOPAVDPDIGPMADQIFYOIFPDRFARSLPREAQDHVYVHHAAGEILIR
HDDEPVTAQAGSSTFYGGDISEKLPYRKUGVLYLNPVKAPASVAKYJYDEYR
HVDPFGDQALRLRHTQOLQMLRLVLDGVFNHSGDSHAMFDRHNRGTGACHNES
PMRWDSFSDGTALRWLGYASLPKLDYQESLVNEIYRGEDSIVRHMLKAPNMGM
RLVVMHLGEAGGRNNMOHVAGITAEKTOPEAYLVGHPGDAROWIQAQVEDAM
NYGTFPRLGFLANDISYDPOIAQCTOMAMNDYRLGSHQOOLRMNOJDSHT
ARKKITLGRDIARPLAVVWLTTPGVPCTIYGEGVGLDGRNDPCKRPPWVEKOD
TALFALYRMTALRKKSQALRHGGCOVLVAEDNVVVFVRVLNQOQVLAVALNRGEACEV
VLASPLANVOWCKRGHQLTDGILALPAISATVWMN"

CDS
misc_feature
gene
CDS

misc_feature 2538. .5551
/note="previously sequenced by another group and can be
found in GenBank Accession Number M37702"
complement (3787. .4368)
/gene="yajB"
complement (3787. .4368)
/gene="yajB"
/codon_start=1
/transl_table=1
/product="hypoetical 14.0 kd protein"
/protein_id="AAB40160.1"
/db_xref="GI:1773088"
/translation="MNFLEMLHMLHAESLSGNLLADFYKGNPESFPDPAVAGITH
HRRIDVLDNLPEPREAREWFRSETRVAPITLDDMMDFLSRMSQSLSPDFOEFV
CYAEQVMTLPPSPRFLNNYLMSEQLVYRMDFIQVTLNGASRRPLDLRL
DSWYDDAHAYDALETRFWCFPRMMAQASKAL"
4461. .5531
/gene="queA"
4461. .5531
/gene="queA"
/codon_start=1
/transl_table=1
/product="S-adenosylmethionine:tRNA
ribosyltransferase-1 isomerase"
/protein_id="AAB40161.1"
/db_xref="GI:1773089"
/translation="MRVTPSPFELPSLTAHYMPERRSGRLSLDGPGLTHGPT
DILDKNPGLVFNNTVTRVRLPFRKASGKTEVLYVEMDDKRLIAIRASKAPK
PGEALLGDDESLNATMTAHGALFEVERDESDVADLISIGHMPLPYITDPDEDA
DRELQYQVYSEKPAVAAPAGLHPEPLEKIRAGVENAFTLVAGACTQFQVRVD
TIEDHIMSHSYAEVPQDVDAVLAARGRVIAVGTTSVRSLSAQAANKNDIEPF
FDTOLEIYFGQYKVDALVTNFHLPESTLIMLVSAFAGYQTNNAKYKAAVEKRF
FSYGDMLFITYNQALNERVE"
5372. .7194
/note="previously sequenced by another group and can be
found in GenBank Accession Number M63939"
5587. .6714
/gene="tgt"
5587. .6714
/gene="tgt"
/codon_start=1
/transl_table=1
/product="transfer RNA-guanine transglycosylase"
/protein_id="AAB40162.1"
/db_xref="GI:1773090"
/translation="MKFELDTDGRARRGLVPDRGVETPCMPYCTGYTKMTPRE
EVBATAGQIILGNTFHLMLRPGQELKHLGDLHDEQMGKPIITDSGGFQVSLGDIR
KTIEQGVHRNPINDGPILDLDEKSMELGYDLSQIVMTFDECTPYPADMDYAKRME
MSLRMAKRRERFDSGNKRALFGIIOGSVEEDLRDISYGLVDIGFDGYAVGLANG
EPKADHRIIEHYCPQIPADKPRYLVGKPEDELVGVRGIDMPCVMPTRBARNGH
LFTVDGVVTRAKYKSDPDREDQCYTCRYSAYLHHLDRCHIELGARLNTIHN
LRYQRLMAGLKAIEGKLESFVDFYQRGREVPPLAND"
complement (5644. .5811)
/note="previously sequenced by another group and can be
found in GenBank Accession Number Z26340"
6006. .10444
/note="previously sequenced by another group and can be
found in GenBank Accession Number X56175"
6426. .7325
/note="previously sequenced by another group and can be
found in GenBank Accession Number S68715"
6737. .7069
/gene="yajC"
6737. .7069
/gene="yajC"
/codon_start=1
/transl_table=1
/product="hypoetical 11.9 kd protein"
/protein_id="AAB40163.1"
/db_xref="GI:1773091"
/translation="MSFRTSDAVATGADAGQSPMSLIMLVFGLIFYMILRPOOK
RKREHKMLDSIAKGEVLTNGLVGRVTKVANGYIAIALNDTTEVVIKDFVAVL
PKGTWML"

gene
CDS
misc_feature
misc_feature
misc_feature
misc_feature
CDS

SOURCE	Escherichia coli O157:H7
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
REFERENCE	Enterobacteriaceae; Escherichia.
AUTHORS	1 (sites)
MEDLINE	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
PUBMED	Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
TITLE	Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
JOURNAL	Sasakawa,C. and Shingawa,H.
AUTHORS	Complete nucleotide sequence of the prophage VT2-Sakai carrying the
Medline	verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
Genet.	derived from the Sakai outbreak
Genes	Genet. Syst. 74 (5), 227-239 (1999)
20198780	
10734605	
2 (sites)	
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,	
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shingawa,H. and	
Hayaishi,T.	
Comparative analysis of the whole set of RNA operons between an	
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an	
Escherichia coli K-12 strain MG1655	
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)	
20557356	
11108008	
3 (sites)	
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,	
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,	
Yamamoto,K., Hayaishi,T., Yasunaga,T., Honda,T., Sasakawa,C. and	
Shingawa,H.	
Complete nucleotide sequence of the prophage VT1-Sakai carrying the	
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli	
O157:H7 strain derived from the Sakai outbreak	
Gene 258 (1-2), 127-139 (2000)	
20564182	
11110050	
4 (sites)	
Hayaishi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,	
Yokoyama,K., Han,C.-G., Ohnubo,E., Nakayama,K., Murata,T.,	
Tanaka,M., Tobo,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,	
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and	
Shingawa,H.	
Complete genome sequence of enterohemorrhagic Escherichia coli	
O157:H7 and genomic comparison with a laboratory strain K-12	
DNA Res. 8 (1), 11-22 (2001)	
21156231	
11258796	
5 (bases 1 to 295741)	
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shingawa,H. and	
Hayaishi,T.	
Direct Submission	
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome	
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,	
Japan (E-mail:kene-gen-info.osaka-u.ac.jp,	
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,	
Fax:81-6-6879-2047)	
genome project.	
location/Qualifiers	
1: .295741	
/organism="Escherichia coli O157:H7"	
/mol_type="genomic DNA"	
/strain="O157:H7"	
/sub_strain="RIMD 0509952"	
/db_xref="taxon:83334"	
131: .709	
/gene="ECs0249"	
131: .709	
/gene="ECs0249"	
/note="similar to GMHA ECOLI gi 1786416 percent identity	
100 in 192 aa (Conserved in E.coli K-12)"	
/codon_start=1	
/evidence=not experimental	
/transl_table=11	
CDS	
gene	
product="phosphotriose isomerase"	
/protein_id="BAB33672.1"	
/db_xref="GI:13359706"	
/translation="MYQDLINELNEAEATLANFLKODANIHAIQRAAVLLADSPFKAG	
GVLSCGNAGSHCDMAHFAEBLGTGYRENRPGPAIAISDVSHISCVNGDFNDIFPS	
RVEAVEGADVGLIGTSGNSANYIKAAAREGMKVITLTGDKGRKAGTAGDIET	
RPHHGVDRIEIHKIYHLIDLIKEVMWK"	
814: .1581	
/gene="ECs0250"	
814: .1581	
/gene="ECs0250"	
/note="similar to YARU ECOLI gi 1786417 percent identity	
99 in 255 aa (Conserved in E.coli K-12)"	
/codon_start=1	
/evidence=not experimental	
/transl_table=11	
/product="putative amidotransferase"	
/protein_id="BAB33673.1"	
/db_xref="GI:13359707"	
/translation="MCELLGSANVPDDICSTFGLVGROGGTGPHKDGKITFYBCK	
GCRTTRDPQPSNPSIALVODPYIKSCSVVAH.RONAKRGVALENHTPTRELWGRN	
WTYAHNGOLDTGYKSLIETNGFRPVGTSEKAFMWLYKLTLQRYRPTGNMAAFKYIA	
SLADELRKGVENMLSDGRVMAVCSTNLHMILTRRAPGVATLLDDVIDESSQTT	
PDVVTVTAIOTPLTNGENTWOKIMPGERMLFGLSRRV"	
complement(1552..2292)	
/gene="ECs0251"	
complement(1552..2292)	
/gene="ECs0251"	
/note="similar to YARK ECOLI gi 1786418 percent identity	
100 in 246 aa (Conserved in E.coli K-12)"	
/codon_start=1	
/evidence=not experimental	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="BAB33674.1"	
/db_xref="GI:13359708"	
/translation="MRKALILAMLLIPCVSFACILLSSSTTPVSKEYKOQLMGSPV	
VIQIRKERITLDLVYKGEYOQLDSTYICKISGLGPKQSGPFSPEGYTSVORNOQ	
LKPDRRYKAINIGPNNAVRAHGEVYKTIHQDCVSIQCYMTNGDIGEIFPVTVGA	
ALVFQPSVOVASIYFRMTDANMRKXSNKDPEWLKPGDYEQTRKRPPTVSVNVA	
GRVYSKPLSHEVOPQLASNYTLPEAK"	
complement(2448..2708)	
/gene="ECs0252"	
complement(2448..2708)	
/gene="ECs0252"	
/note="similar to YAPU ECOLI gi 1786419 percent identity	
98 in 84 aa but differs at N-ter, TTG start (Conserved in	
E.coli K-12)"	
/codon_start=1	
/evidence=not experimental	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="BAB33675.1"	
/db_xref="GI:13359709"	
/translation="MNSGQFSKDYKLAQRKKDNMKIKYLMTLLINNTPLPVAYKDH	
PLQSKWGYRDAAHVHPDWILYKLTDLRLRRGTALFG"	
complement(2727..2987)	
/gene="ECs0253"	
complement(2727..2987)	
/gene="ECs0253"	
/note="similar to DINJ ECOLI gi 1786420 percent identity	
97 in 86 aa (Conserved in E.coli K-12)"	
/codon_start=1	
/evidence=not experimental	
/transl_table=11	
/product="damage-inducible protein DinJ"	
/protein_id="BAB33676.1"	
/db_xref="GI:13359710"	
/translation="MAANAFAVRARIDEDLNQAADVLVAGLITISDLVRIITTKVARE	
KALPDLREPMDLTIOGSKNSEAGVDVHKAKADDDLFDKLGV"	
3197: .3946	
/gene="ECs0254"	
3197: .3946	
CDS	

```

/gene="ECs0254"
/note="similar to YAPL ECOLI gi|1786421 percent identity
98 in 249 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative lipoprotein"
/protein_id="BAB33677.1"
/db_xref="GI:13359711"
/translation="MSLSPIPSFVLSGLILICLPSPSPASTTSHTSPVARMON
RAILKQYQTLKQVSYIVGNAESRAALRQNRQIKPPEPPAPLAKSDRMWA
LAENNHPLSSDHLNTEVALHRLQQLGKPYWGTRPDKGFCDSGLVFAVKNKLE
AKPRTANEMVHYRATVANNDLRRGDLFPFHHSRIADHVMVYLGQGFIESPT
GETIRVSRLEAPFWQDHLGARILTEETII"
4122..4418
/gene="ECs0255"
4122..4418
/gene="ECs0255"
/note="similar to N-terminal part (1-98 in 165 aa) of
YAPL ECOLI gi|1786422 percent identity 93 in 98 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB33678.1"
/db_xref="GI:13359712"
/translation="MSEYRRYIKGTMFTVNLNRSHLLTAQFQKLRAITNVR
DRPEINAMVVLPEHMCITWLPESDDFSSRWEIKKQFHACGLKINMPR"
complement(4764..6503)
/gene="ECs0257"
complement(4764..6503)
/note="similar to FHIR ECOLI gi|1786423 percent identity
98 in 578 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="flagellar biosynthesis"
/protein_id="BAB33679.1"
/db_xref="GI:13359713"
/translation="WLSRSDLLTLTINFIYVVGARISVSARFTLDMPGKQMAL
DADINAGLINOAQOTRKDVASEDFGADGASKFVRGDAIAGMMILAINLIGVC
IGIRKYNLSADAAQQOYVIMTIGDLVAQIPSLISTAAIIVRVSNDGIADHSH
QILASPSVLYATAGIMFVAVPQPHLPFLFSGALGTMRMSKQPAEAEKSL
ETTRTETSEOOVMEETPLIPISLIGKVLAVDKQGNPLQGRIGVQVLS
DGNVILPEIRIRENFRKPSQVAIFINGIADADIPADMLAPSESTGEIDGVL
GNDPAGMPVTWIOPAQAKALNMGQYITDSATITVKNKIVSYIPDLPNYDTIO
LHNRLSSWAPRLAEDLSALNYSQILKRYALTLTEGSLRDIIVITATLVASSAVTD
HLLAADVRLALRSITHPFVRKQELTVYTLNLEMLTNVVAQAQGGKMDSVF

```

Alignment Scores:

```

Pred. No.: 4,73e-34 Length: 295741
Score: 762.50 Matches: 168
Percent Similarity: 61.31% Conservative: 76
Best Local Similarity: 42.21% Mismatches: 129
Query Match: 37.77% Indels: 25
DB: 1 Gaps: 6

```

US-09-889-756a-2 (1-412) x AP002551 (1-295741)

```

Qy 11 ALaAlaAlaLeuAlaAlaAlaValaLeuValaLeuSerSerCySgLyGlyGlyasp 30
Db 270174 GCGGCTTCCTGATCCTCTCAGCAGCTTAGCCCTAACAGATG---GAGCACAACAG 270118
Qy 31 ALaAlaGInGlyGlyGInProAlaGlyArgGluAlaProAlaProAlaValaGlyVala 50
Db 270117 GCCCAACAAGGTGGCCAG-----CAGATGCCCGCCGTTGGCGTAGTA 270076
Qy 51 ThrValHisProGInThrValaLeuThrValaGluLeuProGlyArgLeuGluSerLeu 70
Db 270075 ACAAGCAAAACTGAACCTCTGAGATCAACACGAGCTTCGGGGCGGACCAAGTCCCTAC 270016
Qy 71 ArgThrAlaAspValaArgAlaGlnValaGlyLeuIleGlnIlyArgLeuPheGlnGlu 90

```

```

Db 270015 CGGATCCCAAGAGTTCCTCTCAAGTAGCGGAGATTATCTCGAAGCCTAATTTCAAGAA 269956
Qy 91 GlySerTyValaArgAlaGInGInProLeuTyGInIleAspSerSerTyTyGluAla 110
Db 269955 GGTAGCCACATCGAAGGAGGTCTCTCTATCAGATTAATCTCGGACCTATCAGGCG 269896
Qy 111 AsnLeuGluSerAlaArgAlaGInLeuAlaThrAlaGlnAlaThrLeuAlaTyValaAsp 130
Db 269895 ACATAGCAGAGTGGGAAGGATCTGCGAAGGCCAGGCTGAGCCATATATGCCAA 269836
Qy 131 AlaAspLeuAlaArgTyTySProLeuValaAlaAlaGluAlaValaSerArgGInGluTy 150
Db 269835 TTGACGTGATCGTTATTCAGAACTGCTCGTACATCAGTACATCAGTAAAGCAAGTAC 269776
Qy 151 AspAlaAlaValaThrAlaTyArgSerAlaGluAlaGlyValaTyAlaAlaGlnAlaAla 170
Db 269775 GATCAGGCTCTGCTGATGCGCAACAGCGAATGCTGGGTAACTGCGGGAAGAGCTGCC 269716
Qy 171 IleTySerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190
Db 269715 GTTGAACCTCGCGGATCATCTGCTTACACCAATCACTCTCCGATTAGCGGTCCG 269656
Qy 191 IleGlyGInSerTyValaSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210
Db 269655 ATTGTATGATCGAAGCGTGAAGGCGCATTTGATCAGAACGCTCAGCGACTGCGCTG 269596
Qy 211 AlaThrIleArgGInThrAsnProMetTyValaAsnValaThrGInSerAlaSerGluVal 230
Db 269595 GCAACCGTCGACCACTTGAATCCGATATGATGATGACCAAGTCCACCAACGACTTC 269536
Qy 231 MetTySerLeuArgArgGInIleAlaGluGlyTySerLeuAlaAlaAspGlyValIleAla 250
Db 269535 CTGGCGCTTGAAGCAAGAACTGGGGAATGCAAGCTGAAACAAGAACGGCAAGCCAA 269476
Qy 251 ValGlyIleTySAspAspGlyThrValTyProGluTySgIlyArgLeuPheAla 270
Db 269475 GTGCGCTGATCACCACTGACGCGATTAAGTTCGCGAGACGCTGACGTAATTCCT 269416
Qy 271 AspProValaValaAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValaProAsnAsp 290
Db 269415 GACCTTACCGTATGATCACACCTGAGTCTATCACCTTACCGCTATCTCCGAACCG 269356
Qy 291 GlnAsnIleLeuMetProGlyLeuTyValaArgValaLeuMetAspGlnAlaAlaValaAsp 310
Db 269355 GATCACACTCGTCCGCGGATGTTCTGTCGGGCAAGCTCGGAAGAAGGGCTTAATCCA 269296
Qy 311 AsnAlaPheValaValaProGInGlnAlaValaThrArgGlyAlaTyAsp-----ThrVal 328
Db 269295 AACGCTATTTTAACTCCGCAACAGGGCGTAAACCGTACGCGCGTGCATGCCACCGTA 269236
Qy 329 MetIleValaAsnAlaGInGlyGlyMetGluProArgGluValaThrValaIleGInGln 348
Db 269235 CTGGTGTGTGGCGGATGACCAAGTGAACCCGCTCGATCGTTGCAAGCAGGCTATT 269176
Qy 349 GlyTySerAsnTrpIleValaThrSerGlyLeuTySAspGlyAspTyValaValaGluGly 368
Db 269175 GCGCATAGTGGCTGCTGATGACGAAGGCTCTGAAGCAGGCGATCGCTAGTAATAGCGG 269116
Qy 369 IleSerIleAlaGlyIleThrGlyAlaTySAspValaThrPro----- 382
Db 269115 CTG-----CAGAAAGTGGCTCTGCTGCTCCAGTAAAGCA 269080
Qy 383 LysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGInSerGlyValaGln 400
Db 269079 CAAAGATTACCGCTGATTAATACCAACCAAGCCGCA-----AGCGGTGCTCAG 269032

```

RESULT 21

```

AB016756/c 300817 bp DNA linear BCT 24-DEC-2002
LOCUS Escherichia coli C7073 section 2 of 18 of the complete genome.
DEFINITION AB016756 AB014075
ACCESSION AB016756.1 GI:26106632
VERSION

```

```

SOURCE          keywords
ORGANISM        Escherichia coli CFT073
REFERENCE       Escherichia coli CFT073
AUTHORS         Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
                Enterobacteriaceae; Escherichia.
TITLE           1 (bases 1 to 300817)
JOURNAL         Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P.,
                Raiko, D.A., Buckles, E.L., Lion, S.-R., Boutin, A., Hackett, J.,
                Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C.,
                Perna, N.T., Mobley, H.L.T., Donnenberg, M.S. and Blatterer, F.R.
                Extensive Mosaic Structure Revealed by the Complete Genome Sequence
                of Uropathogenic Escherichia coli
                Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
REFERENCE       12471157
AUTHORS         2 (bases 1 to 300817)
                Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P.,
                Raiko, D.A., Buckles, E.L., Lion, S.-R., Boutin, A., Hackett, J.,
                Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C.,
                Perna, N.T., Mobley, H.L.T., Donnenberg, M.S. and Blatterer, F.R.
                Direct Submission
                Submitted (20-JUN-2002) Genetics Laboratory, University of
                Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES        1. 300817
                location/Qualifiers
SOURCE          /organism="Escherichia coli CFT073"
                /mol_type="genomic DNA"
                /strain="CFT073"
                /db_xref="taxon:199310"
                complement(294. .1079)
                /locus_tag="c0330"
                /locus_tag="c0330"
                /note="Residues 2 to 261 of 261 are 83.58 pct identical to
                residues 1 to 261 of 261 from GenPept.129 :
                >gb|AA122652.1| (AE008876) putative regulatory protein,
                decr family |Salmonella typhimurium LT2]"
                /codon_start=1
                /transl_table=1
                /product="Putative deoxyribose operon repressor"
                /protein_id="AA078818.1"
                /db_xref="GI:26106633"
                /translation="WMEETKQKERIRRLIEILKTDRIHLKDAAMLEVSWVTRRDIAH
                QGEDEPLPLTLGGYIVMVKPAPMPVIOVPRNHRDLPALIAAGVNDDELIPD
                NGBEIPPLIVSMIPALIFGTGICSHRYFVALNEKPVNTALICGTYRARDAPFADAN
                SSPLDNSPKRIFISAGSDVHFGEVSNPEDLATKKRAMRGIRKILLARHALFDEV
                AASLAPLSAPDIVISDRPLPADVYTHCRNAAVAVITPDSDE"
                1382. .2302
                /locus_tag="c0331"
                /locus_tag="c0331"
                1382. .2302
                /locus_tag="c0331"
                /note="Residues 1 to 306 of 306 are 96.07 pct identical to
                residues 1 to 306 of 306 from GenPept.129 :
                >emb|CAC34669.1| (AY087925) unnamed protein product
                (Salmonella typhi]"
                /codon_start=1
                /transl_table=1
                /product="Putative ribokinase"
                /protein_id="AA078819.1"
                /db_xref="GI:26106634"
                /translation="MDIAIVGSNVNDLITYTNQPKEGETLLEAPAKIGCGGKANOQA
                VAAKINSKYLMTKEVGDDIFADNTINILSWGINTLYYKRCSTSSGVAIPFNANS
                NSNLIITKGANKFLSPEDIDRAADELKKCLIVOLEVLETVHAIIEPKKQNTIYL
                LNPAPRELDMSCYAAKCCDFPIAPETLELITGTSVUTYTHIRLAASGLVDKGLNNTI
                VYMSSEKALMTMRDQEVHVPKVNADVTGAGAPFGFSHYVVSGLDEVAALIKKAA
                LFAAFSTYGTGKQSSSYSPISQFNEFTLINE"
                2330. .3646
                /locus_tag="c0332"
                /locus_tag="c0332"
                2330. .3646
                /locus_tag="c0332"
                /note="Residues 1 to 438 of 438 are 96.11 pct identical to
                residues 1 to 438 of 438 from GenPept.129 :
                >gb|AA122650.1| (AE008876) putative L-fucose permease
                [Salmonella typhimurium LT2]"

```

gene
/codon_start=1
/transl_table=11
/product="Putative L-fucose permease"
/protein_id="AAN78820.1"
/db_xref="GI:26106635"
/translation="MNDKNIITGPDGVYKNTPLRPFLLISGLFPLMGAALNIIILT
QKPSVLSNLSALVQSAFTGGYFLIIPSLVTKTSYKVALRLNI SQTPIGAG
PPASHAAITYTFLAALFALIGLSLETAANYTSSMGPKPAVTLRLNI SQTPIGAG
AGSILIKVATSGESLEKQMSNACQILANFKVLMENLEPTKVMIMLVVVMY
FLITLPTKQVATSHHRRPSAMDRLRLANPFRRLI VAOFLVGVQVAVSFTR
LALGDELIVRASNFYVSPACFIFGEIATLNIPTREPKVLIVSICALFLAVY
ALAPSSAAYVAVSVLPGCVATIVAGTIDTQNEHTAGAVIYMAIYGAAVPRA
TGGYIADMFHSIQLSFLVSMLCFTYVGVYFWRSEKVRTALAEVTRAS"
3658.4671
/locus_tag="c0333"
/locus_tag=3658.4671
/locus_tag="c0333"
/note="Residues 1 to 336 of 337 are 83.03 pct identical to
residues 1 to 336 of 337 from GenPept.129 :
>gb|AAJ2649.1| (AE008876) putative cytoplasmic protein
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="Putative cytoplasmic protein"
/protein_id="AAN78821.1"
/db_xref="GI:26106636"
/translation="MSTRINLMRALFGSKPRILLENSDPFVTSFRYSGVEGLKIANSI
RSHLLILPMGCMWDQAPDGHGLTMCMPROPAPATEVIRTGCFAPHSGLANCP
SSEDIHLIFEGACAMDEAWLEIDGDLRLNGREYVYMGCHHYLAOPTVLLKST
LPDIKATYNLASVDMPLQVMGCHMYATIPATSSQNPDEILKRSVSHVNPTRQ
WLAFLNIMQGEASLSTISQPEFYDPELIVFADKLDAITDQPEFRMISPGDTVTFR
YSAELNVTYRTWLIVNGEQVAAPALPATCRPEGYLAARNGTLLIQVAPDQTRPTVT
GIE"
complement(5594.6850)
/locus_tag="c0334"
complement(5594.6850)
/locus_tag="c0334"
/note="Residues 1 to 418 of 418 are 99.76 pct identical to
residues 1 to 418 of 418 from GenPept.129 :
>emb|CAC39291.1| (AJ278144) hypothetical protein
[Escherichia coli]"
/codon_start=1
/transl_table=11
/product="Putative integral membrane protein"
/protein_id="AAN78822.1"
/db_xref="GI:26106637"
/translation="MDFPRILSDVSPRAVLNGLALIGLACKRPEKTCIGTYT
IMGFPIILAKGLVSSLGDDPANIPQAHFGIQQVYPPNBAIVSAQKFGKEMATYF
FAMVNIINIAFTPMKFTFLTGHTLTFSNMVAVILVATGVTLLI VASLVGVAM
VFEPPIAHYPMKYVSGDDVAIGHSTLSYVLACFISFGKNKSHSTEDMNVKSLIF
LRTDVATSPMSIIFLVTCLEFAGADAKELSGKNMFMSIMOSITFAAGVYIILGQ
VRMVAIEIVPAPFKGISDLVGNARPAIDCPVPYPAVNAVILVLSGFAAGLIGMFTL
VLYANLAIIPCVHPHFPGSAAAGVGNATGRRRAIIGAFQSGILITFLVFLPLPUG
DIGPANTTFSADTFCALGILGITVR"
complement(6863.7150)
/locus_tag="c0335"
complement(6863.7150)
/locus_tag="c0335"
/note="Residues 1 to 95 of 95 are 96.84 pct identical to
residues 1 to 95 of 95 from GenPept.129 : >gb|AAK27336.1|
(AF286670) unknown [Escherichia coli]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN78823.1"
/db_xref="GI:26106638"
/translation="MKITVCGNGLSITLMEMNSIKSILKQLSISADVHDVIGSAGK
TPSDIYGTKDIAQLVAOSVAGKIVALENNIDKKMRERLVALTEIGAL"
complement(7166.7609)
/locus_tag="c0336"
complement(7166.7609)
/locus_tag="c0336"
/function="Transport of small molecules"

ACCESSION	AE015072	AE005674	AE015072	AE005630
VERSION				GI:24050630
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PIRNUM				
REFERENCE				
AUTHORS				
TITLE</				

ACCESSION	genome.
AE016979	AE014073
AE016979.1	GI:30040081
KEYWORDS	
SOURCE	
ORGANISM	Shigella flexneri 2a str. 2457T
REFERENCE	Shigella flexneri 2a str. 2457T
AUTHORS	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
TITLE	1 (bases 1 to 292309)
JOURNAL	Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G. II, Rose, D. J., Darling, A., Mau, B., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C. and Blattner, F. R.
REFERENCE	Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T
AUTHORS	Infect. Immun. 71 (5), 2775-2786 (2003)
TITLE	2 (bases 1 to 292309)
JOURNAL	Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G. II, Rose, D. J., Darling, A., Mau, B., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C. and Blattner, F. R.
REFERENCE	Direct Submission
AUTHORS	Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
TITLE	Location/Qualifiers
JOURNAL	1. .292309
FEATURES	/organism="Shigella flexneri 2a str. 2457T"
SOURCE	/mol_type="genomic DNA"
	/strain="2457T"
	/serotype="2a"
	/db_xref="taxon:198215"
	complement(165. .935)
	/gene="yafY"
	/locus_tag="S0290"
	complement(165. .935)
	/gene="yafY"
	/locus_tag="S0290"
	/function="putative enzyme; Not classified"
	/note="residues 1 to 256 of 256 are 98.04 pct identical to residues 1 to 256 of 256 from Escherichia coli K-12 : B0219"
	/codon_start=1
	/transl_table=1
	/product="putative EC 3.5. amidease-type enzyme"
	/protein_id="AAP15815.1"
	/db_xref="GI:30040082"
	/translation="MPSLKITLLDQPLVWVDGPNALHFRDLEGITGRVYLPEMF
	TSGWPAALASSLAODDVVMMWTRKAAQCNALLAGSVALTQESGVRFLIVEGGTV
	HYFDKRLFRNADHLEHYKAGNARVILEMGRMLPLPVCDLRFVWSRNLINVDLAL
	VYVAMPAPRLHMDQLTLARAIENQAVVAGCGRVSGNCHRGDGRVINPQCEIITA
	TADHAKRTIDAEISWVALREYREKFPAMDADDEFRLM"
	1089. .1562
	/gene="ykeE"
	/locus_tag="S0291"
	1089. .1562
	/gene="ykeE"
	/locus_tag="S0291"
	/note="residues 1 to 157 of 157 are 90.44 pct identical to residues 1 to 157 of 157 from Escherichia coli K-12 : B0220"
	/codon_start=1
	/transl_table=1
	/product="hypothetical protein"
	/protein_id="AAP15816.1"
	/db_xref="GI:30040083"
	/translation="MGRISGGMMFKAITVAALVITSAAMDLLTSLAGETTK
	AAFGVQGHKLPAWVGSGTGTPAQCVTLTGDENVQYMSACKPHDCSSQSLIAWMSBK
	SNWGTGFSAIDETSGEKLTLWNLVNLSDIGKTVLPAALTSLLENHDPGFNPK"
	complement(1605. .4085)
	/gene="yafH"
	/locus_tag="S0292"
	complement(1605. .4085)

```

/gene="yafH"
/locus_tag="S0292"
/function="putative enzyme; Not classified"
/note="residues 1 to 826 of 826 are 97.82 pct identical to residues 1 to 826 of 826 from Escherichia coli K-12 : B0221"
```

/codon_start=1

/transl_table=1

/product="putative acyl-CoA dehydrogenase"

/protein_id="AAP15817.1"

/db_xref="GI:30040084"

/translation="MWRRPTSSKGFVMMILSLIAATVVLLGALFYHRVSLFISLSILIA
WMAAGVAGMSAWLVLPDLTIIIDPENAAPRKSMIASPVYRGKAYMPMRSREKKA
IDAQTWMVGDFQKEDPMKKLIHNPQRLFAESDAFLDGVSEACRANDPQTHLEL
ADLPBELLVALEKRFEPAMITKEEYGAEFAVSARVLQLSGVSGILAITYGVNPS
IGRPGLTGTEDEQDKHLYPLARGOEIPCFALTSPAGSDACAIPDTGITVCMEQO
GGGVLMGLTNMKRYITLAPIATVLAGIKFSDELKGADLGICITALPTTPPV
ETGRHPVPNVFPNGPRGKDVFPVIDYIIGGVMGCGWRMLVEGLSVRGITLES
NSTGSKVALAATGAYSHIRPOFKISTGMKEIEPIARIGNAVYWDMAASLTGTGI
NLGEPPVASAIATKYCHTRGOQSITIDAMDITGGGIMLGSNTLNARAYQAAPAIIV
ESANILITSSMMIFQGAKRCHPYVLEEMEAKNNDVNAFDLPEKHIGHVSNSVRSE
LALRTGTSSTPTGDATKRYOHLNRSLANLALDSVSMALLGSGLKREERSARJG
DILSOLVYLAISKRYDEGENEADLPFWHGVDALYOAQAMDNLINLOPNKVNAGA
LNAVTFPGRHVYLARDKLHKAKKILIOVNAATLRSPRGQVLTPEBNPVGLFEER
LVDTVAADPHIRITCKELGRKLPSTRDELAAHNLTKGLDKDXAAILVRAEBRLSS
LNVDFDDEBELATKRVKLPERKRVKAA"

4289..4867

/gene="gmbA"

/locus_tag="S0293"

4289..4867

/gene="gmbA"

/locus_tag="S0293"

/function="enzyme; Surface polysaccharides"

/note="residues 1 to 192 of 192 are 100.00 pct identical to residues 1 to 192 of 192 from Escherichia coli K-12 : B0222"

/codon_start=1

/transl_table=1

/product="phosphoheptose isomerase"

/protein_id="AAP15818.1"

/db_xref="GI:30040085"

/translation="MYODLIREUNEAEATLANLKDADNHAIQRAAVYLASSFRAG
GKVLSGGEGSHCDAMPHEELTGRVRENRPGYPAIASVDSHCVGNDFNDIPS
RVEVAVGEEGVLLGISSTGSANSVIRAIAAAREKMKVITLTGKDGKRAADLEI
RPHFGVADRIOEHIKVIHILOIEKEMVK"

4875..5009

/repeat_region

/note="REP (repetitive extragenic palindromic) element, contains 3 REP sequences"

5073..5840

/gene="yafJ"

/locus_tag="S0294"

5073..5840

/gene="yafJ"

/locus_tag="S0294"

/function="putative enzyme; Not classified"

/note="residues 1 to 255 of 255 are 99.60 pct identical to residues 1 to 255 of 255 from Escherichia coli K-12 : B0223"

/codon_start=1

/transl_table=1

/product="putative amidotransferase"

/protein_id="AAP15819.1"

/db_xref="GI:30040086"

/translation="MCCELLGSAVNPPTDICSFPTGLVORGSGPRPKDGCITPFYEKX
GCRTRKDPDEFNSPIAKLVODYPTRKSQVVAIHROANKREVALENTHPTPTRLKGRN
WTVAHNQGLTVYKSLFTGNRPVGETSSEKAFPCWLHLKIQRYPRTPGNNAAVFXYIA
SIABELEKRKGGFMNLLDGRVYVAVCSTNLHWITRRAPFCVATILLDDVDIEDPSOTT
PNADVVTATAOPLTGNETWOKIMPGEWRLFCLGERVV"

complement(5811..6551)

/gene="yafK"

/locus_tag="S0295"

complement(5811..6551)

/gene="yafK"

/locus_tag="S0295"
 /note="residues 1 to 246 of 246 are 99.59 pct identical to
 residues 1 to 246 of 246 from Escherichia coli K-12 :
 B0224"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP15820.1"
 /db_xref="GI:30040087"
 /translation="MKRIALILAMLLIPCVSFAGLLGSSSSTTPVSKVEYKQOLMGSPV
 YIQIKFEERTLDLYVMGEGQOYLDSYKICKSGGIGPKQROGDFPKSPGFFYSVQNO
 LKPDSEYKAINIGFNPADRAHGYSGKTIAMIGDCVSGCYAMTNGOIEIFQFTVG
 ALVPGPSVQSVSIYPRMTDANKKHKYSNFNDPFWGQIKRGVYFFQTRKRPVSVAN
 GRVYVSKPLSHVEVQPLASNTLPKAK"
 6302..7456
 /gene="yafL"
 /locus_tag="S0296"
 6302..7456
 /gene="yafL"
 /locus_tag="S0296"
 /function="putative membrane, Not classified"
 /note="residues 10 to 184 of 184 are 96.57 pct identical
 to residues 75 to 249 of 249 from Escherichia coli K-12 :
 B0227"
 /codon_start=1
 /transl_table=11
 /product="putative lipoprotein"
 /protein_id="AAP15821.1"
 /db_xref="GI:30040088"
 /translation="MRFAPTQRHNRBOIKOHPMPAPLKLASDRMOALENNHP
 LSSDLNHTTEVAHRIHQOGLKPYWGCTRGDQSGGLVYAVNKKLEALPTPTA
 NEMHYRRATIVANNDLRRGDDLFFHHSREIADHNGVYLDGQFIESTLGTETIRVS
 RLAEFPWQDFLGARRILTEMTL"
 7453..7536
 /note="REP (repetitive extragenic palindromic) element,
 contains 2 REP sequences"
 7632..8129
 /gene="yafM"
 /locus_tag="S0297"
 7632..8129
 /gene="yafM"
 /locus_tag="S0297"
 /note="residues 1 to 164 of 165 are 94.51 pct identical to
 residues 1 to 164 of 165 from Escherichia coli K-12 :
 B0228"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAP15822.1"
 /db_xref="GI:30040089"
 /translation="MSEYRRYYIKGGTWFPTVNLNRRSHLLTTFQTLNNALINVKR
 DRPFINAMVVDPEHHCIMWLPESDDPSRWREIKQFTHACGKNIWQPFWEHA

Db	131853	ACAGTCAAAACGAACTCTCTGACATCAACACGAGCTTCCAGGTGCGACCACTGCTCAC	131794
QY	71	ATGThralAaspValArgAlaGlnValGlyIleIleGlnLysArgLeuPheGlnGlu	90
Db	131793	CGGATCGCAGAAAGTTCGTCCTCAAGTTGCGGGATTTATCTCTAAGCGTAAATTCCAAAGAA	131734
QY	91	GLYSerTYrValArgAlaGlyGlnProLeuTYrGlnIleAspSerSerThyrGluAla	110
Db	131733	GGTACCGACATCGAAGCTGGTGTCTCTCTCTATCAAGATTGATCTCGCGACTTATCGGGG	131674
QY	111	AsnLeuGlnSerAlaArgAlaGlnLeuAlaThralaGlnAlaThrLeuAlaLysAlaAsp	130
Db	131673	ACATACGACAGTGGCGAAGGATGATCTGGGGAAGGCCAGCGCTCGACGCCAATATCGCGCAA	131614
QY	131	AlaAspLeuAlaArgTYrLysProLeuValAlaAlaGluAlaValSerArgGlnGluTYr	150
Db	131613	TTGACGGGAATCGTTATGAGAAATCGCTCGGTACTCGTACATGTAAGCAAGGGTAC	131554
QY	151	AspAlaAlaValAlaThralaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAla	170
Db	131553	GATCAGCGCTCTGGCTGATGCCGACAGAGGGAATGCTCGCGTAACTGCGCGAAAGCTGCC	131494
QY	171	IleLysSerAlaGlyLysAsnLeuAsnArgSerArgIleThralaProIleSerGlyPhe	190
Db	131493	GTTGAAACTCGCGCGGATCATATCTGGCTTACACCAAGTACCTCTCCGATTGAGCGTCCG	131434
QY	191	IleGlyGlnSerLysValSerGluGlyThrLeuLeuAsnAlaGlyAspThrValLeu	210
Db	131433	ATTGGTAGTACGAAGGTACGGAAGGCCCATTTGTACAGAAACGGTCAGCGACTGTGCTG	131374
QY	211	AlaThrIleArgGlnThrasnProMetTYrValAsnValThrGlnSerLysArgGluVal	230
Db	131373	GCAACCGGCGACGAACTTGATCCGATCTGATGTGATGCCAGTCCAGTCCAGCAACGACTTC	131314
QY	231	MetLysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAla	250
Db	131313	CTTCGCCCTGAAACAGAACTGCGGAATGGACGCTGAACAGAAACAGAACGCAAGCCAA	131254
QY	251	ValGlyIleLysPheAspAspGlyThrValTYrProGluLysGlyArgLeuPheAla	270
Db	131253	GTGTGCGTGAATCACCAGTACGCGCATTAAGTTGTCGCCGAGACGGTATACCTGGAATTCCTG	131194
QY	271	AspProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp	290
Db	131193	GACGTTACCGTTGATCAACACACTGGGTCTATCACCTTACGCGCTATCTTCCGAAACCCG	131134
QY	291	GlnAsnIleLeuMetProGlyLysLeuTYrValArgValIleMetAspGlnAlaValAsp	310
Db	131133	GATCACTACTGCTGCTCGCGGATATGTTGTCGTGCGCGACGCTGTGAAGAAAGCGTTATCCA	131074
QY	311	AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysPhe-----ThrVal	328
Db	131073	AACGCTATTATTTAGTCCCGCAACAGGCGGCTAACCCCTACGCGCGGTGATGATCCACCGTA	131014
QY	329	MetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGlnGln	348
Db	131013	CTGTAGTATGGCGCGCATGACAAATGTGAACCCCTCGCATGTTGCAAGCCAGCGCTATT	130954
QY	349	GlyThraAsnTPPIleValThrSerGlyLeuLysAspGlyAspLysValValGlyGly	368
Db	130953	GCGCATAGTGGCTGTGATACAGAAAGTCTGAAAGCAGCGGATCGCGTATTAATTAAGTGGG	130894
QY	369	IleSerIleAlaGlyIleThrGlyAlaLysLysValThrPro-----	382
Db	130893	CTG-----CAGAAAGTGGCGTCTGTGTCTCCAGGTAAGAAAGCA	130838
QY	383	LysGluTPAlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGln	400
Db	130857	CAAGAAATTACCGCTGATATTAATACCAAGCAAGCCGCA-----AGCGGTGCTCAG	130810

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Proteus mirabilis regulatory protein AcrR, membrane fusion protein AcrB, and inner membrane RND family protein AcrB genes, complete cds.	AY061647	AY061647.1	GI:17973457		
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					

BASE COUNT	1497 a	1065 c	1134 g	1585 t
ORIGIN				
Alignment Scores:				
Pred. No.:	4.25e-36	Length:	5281	
Score:	758.50	Matches:	170	
Percent Similarity:	60.71%	Conservative:	71	
Best Local Similarity:	42.82%	Mismatches:	139	
Query Match:	37.57%	Indels:	17	
DB:	1	Gaps:	6	
US-09-889-756A-2 (1-412) x AY061647 (1-5281)				
Qy	12	AlAlaaleuAlaAlaAlaAlaAlaAlaLeuValleuSerSerCyeglylglyAlaaspAla	31	
Db	940	GCTCTGTTAGGCTTTTCAGCAGCTTGGTTCTTCTCGATGTGGCAGCAAAAATCACTCT	999	
Qy	32	AlAglnGlylglyGlnProAlaGlyArgGlyAlaProAlaProAlaValGlyValAlaThr	51	
Db	1000	GCT---GGAGGTCACCT-----CCTGCTCCTGCTAGAGGTGTGTACA	1041	
Qy	52	ValHisProGlnThrValAlaLeuThrValGlnLeuProGlyArgLeuGlnSerLeuArg	71	
Db	1042	TTAAGATGGAAACACCATGACTATACACAGACTTACCTGGCTCTACATCTGTTACGT	1101	
Qy	72	ThrAlaAspValArgAlaGlnValGlyGlyLeuGlnHisArgLeuPheGlnGlnGly	91	
Db	1102	ATCGCAGAGGTTCCCTCCAGGTGGCGGCAATCTTAAACGCAATTACACCGAAGGT	1161	
Qy	92	SerTyValArgAlaGlyGlnProLeuTyGlnLeuAspSerSerThrTyGlyAlaAsn	111	
Db	1162	AGTATATGAGAGCAGGACATCTTTATCCAAATGATCCTGCTATTTTCAAGTACA	1221	
Qy	112	LeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaAlaAspAla	131	
Db	1222	TTTAAACAGTGTCAAGCTGATTATTAAGCAAAAGCGCAATGCTGAATTGCTGCTGTG	1281	
Qy	132	AspLeuAlaArgTyLeuValProLeuValAlaAlaGlnAlaValSerArgGlnGlyTyAsp	151	
Db	1282	ACTGTAGAGCCCTATTAACCTCTACCTGGCCCAATTATGTCAGTAAACAAGATTGTAT	1341	
Qy	152	AlAlaAlaValThrAlaAlaArgSerSerAlaGlyAlaGlyValAlaAlaAlaAla	171	
Db	1342	ACCGCAACATCTCAGTACGCTCAAGCTGTTGCTGCACATTAAGAAGCTGAAGCTACAGTG	1401	
Qy	172	LysSerAlaGlyAlaLeuLeuAsnArgSerArgLeuThrAlaProIleSerGlyPheIle	191	
Db	1402	ACTTAATCCAAAATAATTAATCTTTGAATATATACCAAGTCAACCCACCAATTTTCGGCGCTTCA	1461	
Qy	192	GlyGlnSerLeuValSerGlnGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAla	211	
Db	1462	GGTAAATCAACGGTAAACAGAGGCGCATTTAGTCCCTGCTGCACAAAGTGGCTTTTACA	1521	
Qy	212	ThrIleArgGlnThrAsnProMetTyValAlaAsnValThrGlnSerAlaSerGlyValMet	231	
Db	1522	ACTGTTACGCAAAATCCATCCGATTATATGTCATGTACACAAATCGAAGTGAAGATTATCTG	1581	
Qy	232	LysLeuArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaVal	251	
Db	1582	AAATTTAAATAAATTAATTAAGAGGCGATTATTCGTCAAGACAAAGCAACCGTGTGT	1641	
Qy	252	GlyIleLysSheAspArgGlyThrValTyProGlnLysGlyArgLeuPheAlaAsp	271	
Db	1642	CACCTTAACCTTAACCTTAATGATCAATCCTACGCTCAAAAAGGTCACTTAAGATTTTCTGAT	1701	
Qy	272	ProValValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGln	291	
Db	1702	GTCAGCCGTTATGATATCCACAGGCTCATATACATATGCTGCATTCGACTTAATCTTAA	1761	
Qy	292	AsnIleLeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaAlaAspAsn	311	

Dd		1762	GGTGAATTACTGCGCTGGATAGTTTGTTCGTACAAATAATGAAAAATGGTAATTCGTCAAAAT	1821
Oy		312	AlaPheValProGlnGlnAlaValThrArgGlyAlaLysASP-----ThrValMet	329
Dd		1822	GCTATTATTGATCCCAACAAGCTGTATTGTGTACTCCTCGGCGAAGAACAAACCATG	1881
Oy		330	IleValAsnAlaGlnGlyIleMecGluPheArgGluValThrValAlaGlnGlnGly	349
Dd		1882	GTGCTTAATTAAGAACAACGCTGGTCAAGTGCTACTTAAGAAAGTATCCAACGAGTTGGT	1941
Oy		350	ThraSentPrlLealValThrSerGlyLeuLysAspGlyAspLysValValGluGlyIle	369
Dd		1942	AATAAGTGCATGTAATAGTAGTGTTCAAGTTGATGACCGGTATCGTTCCGGGCTTA	2001
Oy		370	SerIleAlaGlyIleThrGlyAlaLysLysValThrPolySgIU-----	384
Dd		2002	CAAAAAGCA-----CAACCTGAATGAAGGTTTACGCCACAAGAAAGAAATTTGATGCA	2055
Oy		385	TrrAlaSerSerGluAsnGlnAlaAlaAla-----ProGlnSerGlyVal	399
Dd		2056	ACAGCATCAACTGAAAAATCTGAACCACGAGATCCTCAATATAGGAGTC	2106
RESULT 25					
AL646081/c					
LOCUS	AL646081	197050 bp	DNA	linear	BCT 11-APR-2003
DEFINITION	Ralstonia solanacearum GM1100 megaplasmid, complete sequence;				
ACCESSION	segment 6/11.				
VERSION	AL646081 AL646053				
KEYWORDS	AL646081.1 GI:17431274				
SOURCE					
ORGANISM	Ralstonia solanacearum				
	Ralstonia solanacearum				
	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;				
	Burkholderiaceae; Ralstonia.				
REFERENCE					
AUTHORS	1 Salanoubat,M., Genin,S., Attiguenave,F., Gouzy,J., Mangerot,S., Ariat,M., Billault,A., Brottier,P., Camus,C., Cattoico,L., Chandler,M., Choiseau,N., Claudel-Renard,C., Cunneac,S., Demange,N., Gaepin,C., Lavie,M., Moisan,A., Robert,C., Saurin,M., Schlex,T., Siguer,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A. Genome sequence of the plant pathogen Ralstonia solanacearum Nature 415 (6871), 497-502 (2002)				
JOURNAL	Nature 415 (6871), 497-502 (2002)				
MEDLINE	21681879				
PUBMED	11823852				
REFERENCE	2 (bases 1 to 197050)				
AUTHORS	Boucher,C.A.				
TITLE	Direct Submision				
JOURNAL	Submitted (05-DEC-2001) GenomeScope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31336 Castanet-Tolosan Cedex, France, Fondation Jean Dausser-CERPH, 27 rue Juliette Doda, 75010 Paris, France, LMM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, GenomeScope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex. Laboratoire de Génétique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex				
COMMENT	Christian.Boucher@toulouse.inra.fr http://sequence.toulouse.inra.fr/R.solanacearum.html. Location/Qualifiers 1..197050				
FEATURES	/organism="Ralstonia solanacearum" /mol_type="genomic DNA" /strain="GM1100" /db_xref="taxon:305" /plasmid="megaplasmid" complement(120..2123) /gene="RSP0803" /note="synonym: RS01504" complement(120..2123) /gene="RSP0803"				
SOURCE	CDS				

```

/function="miscellaneous: hypothetical/partial homology"/
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABILE SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD17954.1"
/db_xref="GI:17431275"
/db_xref="GOA:O8XRM8"
/db_xref="SPTRMBL:Q8XRM8"
/translation="MKTSTFKUTGTAALVATFAVSIVTSASAASHADLPPIVAAGSSEPL
AGVASAVSYANGLLPPTSIGATPASSGEVVITTLTGALGTDTDLVAAVOSVTGNTPAS
SAVSAGSVISYANVPDGVASTAGGALSICGGSSNALTGA VGVGTVGSVDPSDPG
ALSAGSGSPITSLTGLTNQLPVGMNPAGALLGAVGTGVLGVSVCSSGALLP
VGYNVOVNGVLGSGNPFGALSGVGSYTGA LGSNPNPGALNGVLTGLSIGIPVSOVNS
LTSPPVGSGVGCAGVACGAVAGSAILIGSVNNVNTSVNAGVGVVANVGNA
ACTVIGVIPGVSVSTRTNSNGTTTIQPLAPVITSILTNLTGALPR"
/complement(816..1494)
/note="RSJ37 or RS06124
LONG REPEAT^37 +
evidence=not experimental
complement(943..1611)
/note="RSJ37 or RS06125
LONG REPEAT 37 +
predicted by Homology"
/evidence=not experimental
complement(2852..4954)
/gene="FusA2"
/note="synonym: RS01903, RSp0804"
/complement(2852..4954)
/gene="fusA2"
/function="macromolecule metabolism; macromolecule
synthesis, modification: proteins and peptides -
translation and modification"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE ELONGATION FACTOR G (EF-G PROTEIN)"
/protein_id="CAD17955.1"
/db_xref="GI:17431276"
/db_xref="GOA:O8XRM7"
/db_xref="SWISS-PROT:Q8XRM7"
/translation="MPROTPIERYRNIGISAHIDAKTTTTERTILFYTGWNHKIGEYHA
DDAAIMDMOEBOERGITITSAATHCMRGWGYPHRINI IDTPGHVFTEVESK
MDVALMGAMVDSVGVPOSETVWRANKVKPRIAFVNKMDRVGDAFRVEKNOME
RLKGNPVPVQIPVCAEDHFRRGVULIUKKAIVMDDAOQYKEFTIDI DELBLPAADMW
HDKRWAEALDEADEALEKETLDGEBALTEIKALARRTTAGETIVPMLCGSAPKKXQV
AMLDAVVALESPDI PSI QGHEKDELFEHANDDEFSLARKITMDPVOQLIFP
RYSSGVSVGCTVNVVPVPEKERLGRI ILOMANQRVEIKDVHADIDAAGLKCATGG
DTLCPDNVUILIERMEPEPVI SOAVEPKTDGDCEKMLNRLAODBSF RVVTDEES
SCOTII ISOMGELHELIVDBMKREGRVEATVKGROVAYRETIRVADVDSGF LKOSG
GRGYGHA VTITQPWPCKGYE FVNDIEKGVY PRFPITADV KGRDITNLACYLAGY PIV
DYKVALTYSKHADVUSENAFMRMGASPFKOAMRRAP ILLEPMNAV EVELPEBYMNZ
VYAESRPVSERVVISAKRA"
/complement(5078..5797)
/gene="RSp0805"
/note="synonym: RS01902"
```

CDS

complement (5078. . 5979)

/gene="RSP0805"

/function="miscellaneous; not classified regulator"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"

/protein_id="CAD17956.1"

/db_xref="GI:17431277"

/db_xref="GOA:Q8XRM6"

/db_xref="SPTREMBL:Q8XRM6"

/translation="MDLAPSPPALPGAQAYERELHASSFGALGPIODALIGAAVVR

ATHGDDMLERRGDPSDGIYCVVEGARIGATSAEGRESILAVSPVWFGSIGLDGQ

TIERRARRLLADGIDRLGTRRLVRPODQLALTLASRTVNOVLKDFEARGILL

RLAYGEIELIDSPAGRALARG"

5934. . 6503

/gene="RSP0806"

/note="synonym: RS01901"

5934. . 6503

/gene="RSP0806"

/function="miscellaneous; hypothetical/global homology"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="PROBABLE TRANSMEMBRANE PROTEIN"

/protein_id="CAD17957.1"

/db_xref="GI:17431278"

/db_xref="GOA:Q8XRM5"

/db_xref="SPTREMBL:Q8XRM5"

/translation="MLRPHNPDXETIMRPLSDHLSVYAYHODGRNITATHEPGIPVIVV

VAATFRSPVIGLIDGVAVYRALLIIVAVYFYLRLDVYFGAMIGLIGMAVWGGH

VAASHTVMLSIGLGLFVIGWIVQFVGHVYGRKRAFDVLDLGLVIGPLFLAETVFA

MLRLALDEVASRRARRARAAPGHHAA"

complement (6511. . 6849)

/gene="RSP0807"

/note="synonym: RS01900"

complement (6511. . 6849)

/gene="RSP0807"

/function="small molecule metabolism; energy transfer;

electron transport"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="PROBABLE FERREDOXIN PROTEIN"

/protein_id="CAD17958.1"

/db_xref="GI:17431279"

/db_xref="GOA:Q8XRM4"

/db_xref="SPTREMBL:Q8XRM4"

/translation="MPVVTESGICQKXTDCVAVCPMDCFHAGNPLVIDPDCIOS

ICAPRCPGAHAENDVPADQREFALANAQLSRAIDVRLTQVOPPLADARAAYQVD

KDALITVPA"

complement (7004. . 8746)

/gene="RSP0808"

/note="synonym: RS01899"

complement (7004. . 8746)

/gene="RSP0808"

/function="miscellaneous; hypothetical/partial homology"

gene

CDS


```

Oy 288 ProhnaAspGlnAsnIleuMetProGlyLeuTyValArgValIleuMetAspGlnVal 307
Db 23881 CCCAACCCGAGCGCATCTGCTGCCGCGCATGTATGTGGCGCGCTGCTGCCGGAAGGC 23822
Oy 308 AlaValAspAsnAlaPheValValProGlnGlnAlaValThrArgGlyAla-----Lys 325
Db 23821 GTGAAGACGAGGCGCTCTGTGTCCGAGAGCGCGCGGAGAGACGCGGCAAG 23762
Oy 326 AspThrValMetIleValAsnAlaGlnGlyGlyMetGluProArgGluValThrValAla 345
Db 23761 CCGTTCGCCCTACGTGTGTGGCAGACGCGCAAGCTGCAGCGCGCGCACCTGTGAACGAG 23702
Oy 346 GlnGlnGlnGlyThrAsnTrpIleValThrSerGlyLeuLysAspGlyAspLysValVal 365
Db 23701 CGCACCGGTGGTGAACAGTGGTGTGGCAGCGCGCTGCGCATCGGACCACTGGTGT 23642
Oy 366 ValGlnGlyIleSerIleAlaGlyIleThrGlyAlaLys---LysValThrProLysGlu 384
Db 23641 GTGGAAGGCTGCCCGCTGCC---GTGCCCGCGCGCGGTCAAGACCAACCGCG--- 23591
Oy 385 Trp-----AlaSerSerGluAsnGlnAlaAlaAlaProGlnSerGlyValGlnThr 401
Db 23590 TGGACGCGCAAGACGCGCAACGTCGACCCGCGCGCGCGCGCGCGCTGTGTGGACG 23531
Oy 402 AlaSerGluAlaLysThrAlaSerGluAla 411
Db 23530 GCCGTGCGGCGACGCGCACCGCTGCCGCC 23501

RESULT 26
AL627267/c 240050 bp DNA linear BCT 06-JUN-2002
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome; segment 3/20.
ACCESSION AL627267 AL513382
VERSION AL627267.1 GI:16501740
KEYWORDS
SOURCE
ORGANISM
SALMONELLA enterica subsp. enterica serovar Typhi
SALMONELLA enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 (bases 1 to 240050)
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaiha,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holtz,E.,
Jagels,K., Krogan,A., Larsen,T.S., Leachter,S., Moule,S., O'Gea,P.,
Pavry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrett,I.B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6858), 848-852 (2001)

JOURNAL
MEDLINE 21534947
PUBMED 11677608
2 (bases 1 to 240050)
Parkhill,J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
Sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES
source
1. 240050
/location/Qualifiers
1. 240050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
213. 473
gene

```

```

CDS
/gene="STY0512"
/note="synonyms: rpmE2, ykGM"
213. 473
/gene="STY0512"
/note="Similar to Escherichia coli hypothetical 9.9 kDa
protein in int-1 each intergenic region ykGM SW:YKGM_ECOLI
(p71302) (87 aa) fasta scores: E(): 1.6e-27, 74.4% id in
86 aa, and to Listeria monocytogenes 508 ribosomal protein
131 rpmE SW:RL31_LISMO (Q9ZH28) (81 aa) fasta scores: E():
7.9e-12, 48.1% id in 79 aa, and to Borrelia burgdorferi
508 ribosomal protein 131 rpmE or db0229 SW:RL31_BORBU
(O51247) (81 aa) fasta scores: E(): 7.8e-11, 43.6% id in
78 aa
Fasta hit to RL31_ECOLI (70 aa), 33% identity in 84 aa
overlap
Orthologue of E. coli YKGM_ECOLI: Fasta hit to YKGM_ECOLI
(87 aa), 74% identity in 86 aa overlap"
/codon_start=1
/transl_table=1
/product="putative 50S ribosomal protein L31 (second
copy)"
/protein_id="CAD04954.1"
/db_xref="GI:16501741"
/db_xref="SPREMBL:08XGM1"
/translation="MKPDHPVYRTVVFHDTSANEVYKVGSTIKTEREILDGVTPY
VTIDVSKSHPFYTGKRTFDSSESAARFQKRFHFGAKRG"
213. 452
/gene="STY0512"
/note="Plam match to entry PF01197 Ribosomal L31,
Ribosomal protein L31, score 48.40, E-value 1.6e-10"
479. 619
/gene="STY0513"
/note="synonym: rpmJ2"
479. 619
/gene="STY0513"
/note="Similar to Guillardia theta chloroplast 50S
ribosomal protein 136 rp136 SW:RX36_GUIRH (P28528; O46911)
(48 aa) fasta scores: E(): 1.6e-09, 56.5% id in 46 aa, and
to Rickettsia prowazekii 50S ribosomal protein 136 rpmJ or
rp456 SW:RL36_RICPR (Q9ZD87) (41 aa) fasta scores: E():
4.9e-08, 56.1% id in 41 aa"
/codon_start=1
/transl_table=1
/product="putative 50S ribosomal protein L36 (second
copy)"
/protein_id="CAD04955.1"
/db_xref="GI:16501742"
/db_xref="SPREMBL:08XF38"
/translation="MVLNLSRNAKQRHPDQIVKRGRLVYICKTNPRFKAQGRKK
RR"
479. 601
/gene="STY0513"
/note="Plam match to entry PF00444 Ribosomal L36,
Ribosomal protein L36, score 23.10, E-value 2.2e-06"
complement (675. 1145)
/gene="STY0514"
complement (675. 1145)
/gene="STY0514"
/note="Orthologue of E. coli YLAC_ECOLI: Fasta hit to
YLAC_ECOLI (169 aa), 81% identity in 155 aa overlap.
Contains two potentially membrane spanning domains"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="CAD04956.1"
/db_xref="GI:16501743"
/db_xref="SPREMBL:Q8Z8U0"
/translation="MTEIORLSEITDIDLVNREKRDNRPFSSIFIKHGLFIAMYA
AMFAATLVMLOSSETIGSWLVLVFLFAFNGFPFDPAPAHYNDIDVLRCVNGE
WYNTRFVPTLITFILOSPOVDNENHKVQLQKMARKEGSLFYDIFTLARAABAR"
complement (1263. 1814)
/gene="STY0515"
/note="synonym: maA"

```

```

CDS      complement(1263..1814)
         /gene="STY0515"
         /EC_number="2.3.1.79"
         /note="Similar to Escherichia coli maltose
         /O-acetyltransferase MAa SW:MAA_ECOLI (P77791) (182 aa)
         fasta scores: E(): 0, 79.7% id in 182 aa, and to Rhizobium
         melioli nodulation protein 1 modL SW:MODL_RHIME (P28266)
         (183 aa) fasta scores: E(): 4.6e-30, 47.5% id in 177 aa,
         and to Escherichia coli galactoside O-acetyltransferase
         SW:THGA_ECOLI () (203 aa) fasta scores: E(): 1.4e-28,
         40.8% id in 179 aa
         Fasta hit to THGA_ECOLI (203 aa), 41% identity in 179 aa
         overlap
         Orthologue of E. coli MAa_ECOLI; Fasta hit to MAa_ECOLI
         (182 aa), 80% identity in 182 aa overlap"
         /codon_start=-1
         /transl_table=11
         /product="maltose O-acetyltransferase"
         /protein_id="CAD04957.1"
         /db_xref="GI:16501744"
         /db_xref="SPTRMBL:O82879"
         /translation="MSDEKQKMIAGALVCPPTDETLCDRLRAPOLIHQVYTTTDELTN
         KQAVILYTRTHPLDVAERNSEGLKSPVITGNINNVIGRAVAVNGVITGDNVVASCA
         VVTXNVPPDVVVGSPNARIKKL"
         complement(1317..1403)
         /gene="STY0515"
         /note="PE00101 Hexapeptide-repeat containing-transferases
         signature"
         complement(1323..1376)
         /gene="STY0515"
         /note="Pfam match to entry PF00132 hexaped, Bacterial
         transferase hexapeptide (four repeats), score 16.40,
         E-value 0.68"
         complement(1377..1430)
         /gene="STY0515"
         /note="Pfam match to entry PF00132 hexaped, Bacterial
         transferase hexapeptide (four repeats), score 13.20,
         E-value 6.2"
         complement(1485..1538)
         /gene="STY0515"
         /note="Pfam match to entry PF00132 hexaped, Bacterial
         transferase hexapeptide (four repeats), score 17.30,
         E-value 0.37"
         complement(1993..2211)
         /gene="STY0516"
         /note="synonym: hha"
         complement(1993..2211)
         /gene="STY0516"
         /note="Fasta hit to YDGT_ECOLI (71 aa), 38% identity in 71
         aa overlap
         Orthologue of E. coli hha (HHA_ECOLI); Fasta hit to
         HHA_ECOLI (72 aa), 99% identity in 72 aa overlap"
         /codon_start=1
         /transl_table=11
         /product="haemolysin expression modulating protein"
         /protein_id="CAD04958.1"
         /db_xref="GI:16501745"
         /db_xref="SPTRMBL:Q9L654"
         /translation="MSDKPLTKTDYLMKLRRCQTITLRIYIEKKYKELSDNLEAVFY
         SAADHRLAELTMNKKYKIPSPVWKFIR"
         complement(2239..2613)
         /gene="STY0517"
         /note="synonym: ybaJ"
         complement(2239..2613)
         /gene="STY0517"
         /note="Orthologue of E. coli ybaJ (YBAJ_ECOLI); Fasta hit
         to YBAJ_ECOLI (124 aa), 95% identity in 124 aa overlap"
         /codon_start=1
         /transl_table=11
         /product="conserved hypothetical protein"
         /protein_id="CAD04959.1"
         /db_xref="GI:16501746"

```

	gene	/db_xref="SPTREMBL:Q9XKX5" /translation="WDEYSPKHDIQAQIKPLCELTLYHDCLANLEESNHGANDPISAV NLQLELIETHIALNALNKIKYNEDNKKLIAIDELVDTFMLFSSYGINTDLDKMR SGNRILFRCFVATAMNPVSLSC" complement (3109 . 6258) /gene="STY0519" /note="synonym: acrb" complement (3109 . 6258) /gene="STY0519" /note="Pasta hit to ACRF_ECOLI (1034 aa), 78% identity in 1033 aa overlap"
CDS		
	Alignment Scores:	
	Pred. No.: 8,81e-34	Length: 240050
	Score: 756.00	Matches: 166
	Percent Similarity: 60.70%	Conservative: 78
	Best Local Similarity: 41.29%	Mismatches: 128
	Query Match: 37.44%	Indels: 30
DB:	1	Gaps: 5
US-09-889-756A-2 (1-412) x AL627267 (1-240050)		
Oy	11 AlaAlaAlaLeuAlaAlaAlaValAlaLeuValLeuSerSerCysGlyLyseGlyValAsp 30	
Dd	7444 GCGGTCTTCGAGTCTCTCAAGCCACTTACGGCTAACAGGATGACGACAAACAGAC 7385	
Oy	31 AlaAlaGInglYgLYGLnProAlaGlYArGluAlaProAlaProValAlaIGlYValVal 50	
Dd	7384 ---CAGCAAAGCGGCCAG-----CAGATGCCAGAAGTTGGGGTTCTC 7346	
Oy	51 ThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGluSerLeu 70	
Dd	7345 ACACTAATAAACCGAACCACTCAATATCACACTGAACCTTCGGGTCGTACCGTTCAC 7286	
Oy	71 ArgThrAlaAspValArgAlaGlnValGlyLytleileGlnValArgLeuPheGlnGlu 90	
Dd	7285 CGTATCGCCGAAGTTCGCCCGCAGGTAAAGCGCATTTATCTGAAGCCGTAAATTGTTGAG 7226	
Oy	91 GlySerTYrValARgaLaGlyGLnProleuTYrGlnleAspSerSerThTYrGluAla 110	
Dd	7225 GGAAATGATATCGAAGCGGGAGCTCTCTCTATAGATTGATCTCGGCATTCACAGGCG 7166	
Oy	111 AsnLeuGluSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAsp 130	
Dd	7165 ACTTAGCAGCAGCGCTAAGGCGATCTGGCAAAAACGCGAGGCCCGCGAATATCGCTGAA 7106	
Oy	131 AlaAspleuAlaArgTYrLYsPProleuValAlaAlaGluAlaValSerArgGlnGluTYr 150	
Dd	7105 CTGACGGTGGAAGCGTTTATCAAAACCTCTGGGTACGCGAATACAGTAAAGCAGGAATAC 7046	
Oy	151 AspAlaAlaValAlaThrAlaLYsArgSerAlaGluAlaGlyValLYsAlaAlaGlnAlaAla 170	
Dd	7045 GATCAGGGGCTGGCTGACGCCGACAACAGCATCCCCCGTTGTCCGACGAAAAAGCCGCC 6986	
Oy	171 IleLYsserAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPhe 190	
Dd	6985 GTTGAACCCGACCGTATCAACCTGGCGGTATACCAAATCACTCACTGCGATTACGGTCGT 6926	
Oy	191 IleglyGlnSerLYsValSerGluGlyThrLeuLeuAsnAlaGlyAspThrThrValLeu 210	
Dd	6925 ATTGTGAAGTGTGTCCTGTAACGGAAGGCGCGCTGTACAGAACCGTCAAGCGTCGCGCTGT 6866	
Oy	211 AlaThrIleLeuArgGlnThrAsnProMetTYrValAsnAlaThrGlnSerAlaSerGluVal 230	
Dd	6865 GCGACAGGCGACGCTGACCCCTATTATTATGTCATGTGACCCAGTCCACCAATGACTTC 6806	
Oy	231 MetLYseuArgArgGlnIleAlaGluGlyLYsLeuLeuAlaAlaAspGlyValIleAla 250	
Dd	6805 CTGGCGCTGAAGCAGCAGCTGGCAAAAGTTCGTGTAAGCAGGAAAAAGCGCAAGCAAG 6746	
Oy	251 ValGlyIleLYsseAspAspGlyThrValTYrProGluLYsGlyArgLeuLeuPheAla 270	
Dd	6745 GTTCATCTGGACAAACCGGATATCAATTAATTCGCGGATGCGGTACGCTTGAATCTCC 6686	

Oy		271	aspprovalValasngIuseThrglInleThreIuearGalalavalProasnaasp	290
Dd		6685	GACGTGACCCTGGTGAACAAGCACCgggTCtATTACTTTGGCGGCATCTCCCTPAACCG	66286
Oy		291	GlAsnilleuMetProglYleuTyxValArGallleuMetAspGlnValAlavalAasp	310
Dd		6625	GATCACACCTTATTGCCAGAAGATGTCCTGGCCAGCTCTGCAAGAGGACAAACCG	65666
Oy		311	ASnalapeIvealValProgInGlnAlavalThrarglYalAsp-----ThrVal	328
Dd		6565	ACGGATTACTCGTGTCACAACAGGCGCTTACCCTACTCCAACGGCGGATGCCACGGTG	65066
Oy		329	MeIIeValasnlaGlnglylGymetGluProArgGluValThrValAlaglInGln	348
Dd		6505	CTGGCGGTGGCGCGCTGATACAAAGTGGAAACCCGCAATTCGTGCGAACGCAAGCAGGCATC	64466
Oy		349	GIYThraenTPPIleValThrSerGylLeuysAspGIyaSpLyeValValValJclugly	368
Dd		6445	GGCGATAAGTGGCTGGTCACTGACCGGTGTGAAGCGGCGACCGCGTGTGTGACGCGG	63866
Oy		369	IleSerllealaglyIleThnglyAlaIaylsVsValThrPro-----	382
Dd		6385	CTG-----CAAAAAGTACGTCCTCGGCGCCACAGTTAAAGTC	63566
Oy		383	-----LySGlUTrPaIaseSerGluasnGlnAlaalAalABro	395
Dd		6349	CAGGAATTTACCGCGGATACAAACAGCAAGACCGGACCGGATCAACCTGCTCAGCCC	62906
Oy		396	Glnser 397	
Dd		6289	AGGTCT 6284	

```

RESULT 27
AE016842
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL
TITLE
SOURCE
1. 300029
/organism="Salmonella enterica subsp. enterica serovar Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
complement (70 .. 1524)
/gene="ybdw"
/locus_tag="t2339"
/note="Corresponds to STY0569 from Accession AU513382:
Salmonella typhi CT18"

```

gene /pseudo
complement (1620. .2867)
/locus_tag="t2340"
complement (1620. .2867)
/locus_tag="t2340"
/note="corresponds to STY0568 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="probable metabolite transport protein"
/protein_id="AA069933.1"
/db_xref="GI:29138363"
/translation="MSERFKSGIDYVVKQIVVMSIGVAVIMIVRYTLVPIYPEIKQASTL
DNVSAEETIGASFTFFAYCCMQIPCGILNDKPEQAKIMLAGFLITIGTLICAKANG
LAMITGSLMAGGGCASFSSAYSLSSANVFOARALANATINSGA1IMG1GILGSS
VLVKMSAMONVLYIVAAIIVIMCVFTLVIRKASDGAIOAKQOTYTEEKRAP
LESGLGCVPEYFECTCYGLVLIYTMVPSYQOTERFGDGA1GLASALVAVGPGAL
PESHLSDEFRNSKVIILGLEIVAAAMAFVLISPTMTMLVSLTIVGLGIRKAVDPTI
LISPSDEASAKSKIGRAPSLFNFGMSAVAVAPLTGPIIDVYTSKSIISVISACLVV
TGLTIFAATVLPKKATORTASA"
complement (2933. .3811)
/gene="garR"
/locus_tag="t2341"
complement (2933. .3811)
/gene="garR"
/locus_tag="t2341"
/note="corresponds to STY0567 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="2-hydroxy-3-oxopropionate reductase"
/protein_id="AA069934.1"
/db_xref="GI:29138364"
/translation="MKLGFIGLIGMSPMALINLARGHQLHTVTTIGVADSLIGAV
NTEATPDEYFADIFIMVPTPOVEEVLFGESCAKTSLOGKTIYVMSISIPGTRK
FAORINEMDAPVLAIPVAGGEGIGAREGLISIMWGEOKVPRVLPLEDDIGKNTIVG
GNGDQGTCKVANOITVAINITAVSEALVPASKAADVRRQALMGSPASSRILEVNG
EMNIRTEPEGKIALHOKDNLALLOSAKALANLPTATCQELFNTCAANGSSQLDH
SAMVOLALIMNHKLS"
complement (3912. .4688)
/gene="hyi"
/locus_tag="t2342"
complement (3912. .4688)
/gene="hyi"
/locus_tag="t2342"
/note="corresponds to STY0566 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="hydroxypropylate isomerase"
/protein_id="AA069935.1"
/db_xref="GI:29138365"
/translation="MLRPSANLSMLFTPEYDFLERDKAAGSPFGEVMEFYDDIVDVE
NCLVKTSGTSRATETIYDTLVENIRPYANMLAKEDILILEPIHFPMPGRHLGTGTO
ALVILKIDGSONIKIQYDIYHQRMEGELTQTMAMADKIGHLQIDUNPRGEPGTSEE
INVDPIFNVIKQSDVDGWCVEYKPYLTAAGLSWINOYR"
complement (4701. .6482)
/gene="gcl"
/locus_tag="t2343"
complement (4701. .6482)
/locus_tag="t2343"
/note="corresponds to STY0565 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="glyoxylate carboxylase"
/protein_id="AA069936.1"
/db_xref="GI:29138366"
/translation="MAKMRVADAMVYKSGITTAFCVPGCAINPFYSAMRKGGIR
HILAHVHSGASMGYTRAAAGINVCULGTSGAGMDITALYSASADSIPIICITG
QAPRLHKEFOAVDIEAIKPVSKAAVYREALVPRVLOQAFHLMRSCRPPVAVL


```

Db      51367 CTGGCGCTGAAGCAGAGAGCTGGCAAAATGTTGCTGTAAGCAGAAAAACGCGAAACGGAAG 51426
Qy      251 ValGlyIleLysPheAspAspGlyThrValTyrProGlyIysGlyArgLeuLeuPheAla 270
Db      51427 GTCGATCTGGTGACAGAGGAGGATATCAAAATTCGCCGACGTCCGATACGCTTGAATTTCTCC 51486
Qy      271 AspProValValAangLysSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAsp 290
Db      51487 GACGTGACCGTTGACCAAGACCGGGGCTCTATTACTTGGCCGACATCTTCCCTAACCCG 51546
Qy      291 GlnAsnIleLeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAsp 310
Db      51547 GATCAACCTTATATGCGCAGGAATGTTCTGTGCGCAGCGCTGACGAGGAGGACAAACCG 51606
Qy      311 AsnAlaPheValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrVal 328
Db      51607 ACGGCACTTACTGTTCCCAACAGGGCGCTTACCGCTACTCCACGGCGGATGCCAGGTG 51666
Qy      329 MetIleValAenAlaGlnGlyGlyMetGluProArgGluValThrValAlaGlnGln 348
Db      51667 CTGGGTGGTGGCGCTGATTAACAAAGTGAAACCGCCAAATCGTCGCAAGCCAGGCGATC 51726
Qy      349 GlyThrAsnTyrPleValThrSerGlyLeuLysAspGlyAspLysValValAlaGluGly 368
Db      51727 GGCATTAAGTGGCTGCTGACTGACGAGGTTGAAAGCGGCGACCGCTAGTCTCAGCGGG 51786
Qy      369 IleSerIleAlaGlyIleThrGlyAlaLysLysValThrPro----- 382
Db      51787 CTG-----CAAAAAGTACGTCCTTGGCGCACAGCTTAAATG 51822
Qy      383 -----LysGluTyrAlaSerSerGluAsnGlnAlaAlaPro 395
Db      51823 CAGAAATTACCGCGGATTAACAAACAGCAGCCGATGATCAACCTGCTCAGGCC 51882
Qy      396 GlnSer 397
Db      51883 AGGTCT 51888

RESULT 28
PSENVCD      5600 bp      DNA      linear      BCT 16-AUG-1995
LOCUS      PSENVCD      Pseudomonas aeruginosa mexa and mexb genes, complete cds and outer
DEFINITION      L11616
ACCESSION      L11616.1 GI:438852
VERSION      homologous.
KEYWORDS      Pseudomonas aeruginosa
SOURCE      Pseudomonas aeruginosa
ORGANISM      Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
              Pseudomonadaceae; Pseudomonas.
REFERENCE      1 (bases 1 to 5600)
AUTHORS      Poole, K., Heinrichs, D.E. and Neshat, S.
TITLE      Cloning and sequence analysis of an EnvCD homologue in Pseudomonas
              aeruginosa: regulation by iron and possible involvement in the
              secretion of the siderophore pyoverdine
JOURNAL      Mol. Microbiol. 10 (3), 529-544 (1993)
MEDLINE      95058196
PUBMED      7968531
REFERENCE      2 (bases 1 to 5600)
AUTHORS      Poole, K.
TITLE      Direct Submission
JOURNAL      Submitted (02-JAN-1994) Keith Poole, Microbiology and Immunology,
              Queen's University, Kingston, Ontario, Canada
COMMENT      Original source text: Pseudomonas aeruginosa (strain CD10) DNA.
FEATURES
  source
    1..5600
      /organism="Pseudomonas aeruginosa"
      /mol_type="genomic DNA"
      /strain="CD10"
      /db_xref="taxon:287"
      403..1554
        /gene="mexA"
      403..1554
        /gene="mexA"
  CDS
    gene
      403..1554
        /gene="mexA"
    CDS
      403..1554
        /gene="mexA"

```

```

/gene="mexA"
/function="possible function in pyoverdine secretion"
/notes="envc homologue; ORF A"
/codon_start=1
/evidence=experimental
/transl_table=11
/protein_id="AA074436.1"
/db_xref="GI:438853"
/translaction="MORTPAMRVLVLPALVAISALSGCGKSEAPPAQTPPEVITLE
AOTVINTLPGRTNAPRIAEVRPONGIILKLFPKESDPKAGQOOLYQIDPATYEA
YSAONMLASTQSOAORXYKLVDADQVSKQVADANAAYOSKAAVQEARINLRITY
LSPIGRIGHSATBGLVTNGQANMATYQOQDPIYVDTQSTALIRRLRLASQ
LERAGDAARVSLKLEDGSOYPLEGRLESEVSDGTSVTRAAFPNNBELPEM
FVAOLQEGVQKQAILAPQOGVTRDLQGAATLVAANQKVELRVLKADRVIDKMLV
TEGLNMGDKLITTEGLQFVQPGVEKTVPAKNVASAKADAPAKTDSK"
1570..4710
/gene="mexB"
1570..4710
/gene="mexB"
/function="possible role in pyoverdine secretion"
/notes="envd homologue; ORF B"
/codon_start=1
/evidence=experimental
/transl_table=11
/protein_id="AA074437.1"
/db_xref="GI:438854"
/translaction="MSKFFIDRPIFAMVILVIMLAGLSILSPVNOYPAIPAPA
VOVSYGSAEYTDQTVVOVIEBOONQINDLRVISESNDSGMTTVEEGQDPDI
AOVOYONKLOLAPLPLPOBYORGIYTRAVKRFPMVWCVGSDGSMTEKEDSNYIS
NIODPLSRKGVDFVQFSQYSMRTMLDPAKNSYQLTFGDYSALQONVOISQO
LGLPAPVKGQQLNAITIGKTRLOTAQFENILUKNPDGQVSLKQVADVGLQGDYS
INAQFNGSPASGIAIKLATCANALDPAKALRQITANLEPMPQGMKRVVYDTPVVS
ASJHEVKTIGBALIVFLVMYLFQNFATLPIPTLAPVVLGTGFLAEGFSINT
LTMGEVATIGLVDLAIYVENVERVMAEGLSPREAPKSMGOIGALVAGLAVIS
KAFIPMAFRGSGTVGYROSTITVAMLSVVALIIPALCATMILKPIEKDGHGB
KGFPRGNMELSTHGTGREGVASILKRAPIPLIYVIVAGIMKFRIPTRALPD
EDGVLPFAOVQTPGSSAERTQVVDMSREYLLKESSSVSFVTVGFNPAQGSS
GMAFIMLKPEWERPGENSVFELAKRQNMFFSKDAMVFAAPSVLDELGNATGFL
FLDOAGVGHVLLQARKFELMAQNPAQYRPNQMSDEPOYKLEIDDEKASALCV
SLADINVSIAWSSVYNDPIDRGYKRYLQYRPPARNPDLISKVYLVNPKGEVY
PNNAPATGKVEGSPKLERYNCPAMEIIGEPAPIGLSGGMAAVEITVQIPGVGY
SWTGLSIEERLSQAPALYALSLVFLCLALYESWSIPESVMLVVPVIGVALLA
TSNRGLSNDVFPQVGLTTTIGLAKNALIVEPAKLEHGKGIVEAIAEACNRRLP
IWNLSLAFLIGVPLAISTGAGSGSHALIGTVIGKWTATVLAIFWVPLFYAVSTL
FXDBASKQASVEKQ"
4712..5600
/gene="oprM"
4712..5600
/gene="oprM"
/notes="ORF C; formerly called oprX"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AA074438.1"
/db_xref="GI:833173"
/translaction="MKRSFSLAAVAAVVLSGCSLIPDYQRPAPVAAYPOGAYGON
TGAAYPAADIGWREFRPDLOOLQIGVALENNRDLRVALNTEPAOYRIORADLF
PRIGVSGSGRPLRPGDLSITGSPAISQVGLTGTAWLDELGRARSRLDQALDY
LATBOARSQOTLVASVATATYITUKADQLOLQDTYLTQYKSPULTRSDYGA
SAULDRQAQTVAGARATTAQYTRLVAAQDONALVLLSGIPANLPGLDLDOTLTE
VPALPESDLQRRPDLIEAE"
BASE COUNT      1004 a      1845 c      1796 g      955 t
ORIGIN
1
Alignment Scores:
Pred. No.:      9,79e-36      Length:      5600
Score:          753.00      Matches:      174
Percent Similarity: 58.60%      Conservative: 68
Best local Similarity: 42.13%      Mismatches: 125
Query Match:    37.30%      Indels:      46
DB:             1      Gaps:      10
US-09-889-756A-2 (1-412) x PSENVCD (1-5600)

```

[illegible]

```

Oy 360 AsgylAspLysValIValValGluGlyIleSerIleala-----GlyIleThrGlyIala 377
Db 1420 GCCGCGCAACAAATCATTATCCGAAAGCGCTGCAGTTCCGTCAGCCCGGGGTGTGAG---CTG 1478
Oy 378 LysLysValThrProLysGluTrpAlaSerSerGluAsnGlnAlaAlaAlaProGlnSer 397
Db 1477 AAGACCGTGGCGGGGAAAGAAATGTGCGTCCGCGCAGAAAGCGGACGCCGCTCCG----- 1530
Oy 398 GlyValGlnThrAlaSerGluAlaLysThrZlaSerGlu 410
Db 1531 -----GCCAAAAACGACACGCAAG 1548

```

RESULT	29				
AE004479					
LOCUS	AE004479	13887 bp	DNA	linear	BCT 19-FEB-2003
DEFINITION	Pseudomonas aeruginosa PA01, section 40 of 529 of the complete genome.				
ACCESSION	AE004479	AE004091			
VERSION	AE004479.1	GI:3946272			
KEYWORDS					
SOURCE					
ORGANISM	Pseudomonas aeruginosa PA01 Pseudomonas aeruginosa PA01				

REFERENCE
1 (bases 1 to 13987)
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,

TITLE
Complete sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
Nature 406 (6799), 959-964 (2000)

MEDLINE 20437337
PUBMED 10984043
REFERENCE 2 (bases 1 to 13987)
AUTHORS Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hudnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reiter, J., Sater, M.H., Hancock, R.E.W., Lofy, S. and Olson, M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 13987)

CONSRM TITLE JOURNAL	COMMENT
<i>Pseudomonas aeruginosa</i> Community Annotation Project. (PseudoCAP) Direct Submission Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada	

This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from PseudoCAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.

Class 1: Function experimentally demonstrated in *P. aeruginosa*.

Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).

Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.

Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

FEATURES	Location/Qualifiers
source	1..13987 /organism="Pseudomonas aeruginosa PAO1" /mol_type="genomic DNA" /strain="PAO1" /db_xref="taxon:208964"
gene	69..680 /locus_tag="PA0417" /note="synonym: cnpE"
CDS	69..680 /locus_tag="PA0417" /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)" /codon_start=1 /transl_table=1 /product="probable chemotaxis protein" /protein_id="AAG03806.1" /db_xref="GI:9946273"
gene	/translation="MLAIFLAALLFGFAFNVSPGAVSETLRGLTGGRPALIVQGS SLIDAVWALLGLTGLLGEQVRIPLTACAAVLAIVGQIRDAWSPPLAEDA GEOGRNAGAGAAISLSPKVVYWGALSGIAGIVGTQNOQSLVPFAGFMLSII WCCCALVDMLRNTSLFMHRSVYAGGVLLGLAGALRGL"
CDS	complement(688..2103) /locus_tag="PA0418" complement(688..2103) /locus_tag="PA0418" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAG03807.1" /db_xref="GI:9946274"
gene	/translation="MNTYAQPTSORIARMDPVYFDKIRLPDVLGVLRDYARFPO LASALETLERLDDLPEDFLPALGOLESGLALVERLDEDSLNIYVGAHPSOR LIDALAAQGHNAARTROENAEGRALPAAPAGGLPASAYICLEHAHAPSLA RLPEDSDLDLDRQWPLQRLPDGAETQGLACRASSADGLHAWMTSGSPYAW LHRSRDLASTPELPPLPAQSLRQRTPELAIGLADGLLADRLGFLPYFGQ GNDPRLFEVPOARSIEHPAVFTTPDQVCLLSRGQFEMREARLQPLPFGVD LPADAPOTSAAPGTIVLRENTCEGLDSCETRSISGLAGAPVIXLQALDGM LLAWLEPHRSRDLAQLMHESGRALRIRYGLDLBGLQHMALPDGRIVGCHAH VDLGTFSLRQLRRRLAFA"
CDS	complement(2138..2860) /locus_tag="PA0419" /note="synonym: Y99J" complement(2138..2860) /locus_tag="PA0419" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="AAG03808.1" /db_xref="GI:9946275"
gene	/translation="NRLSRFFVDDPLSIGODLPEAOAHYIGRVLRHATGAVOLFQD SGREIVGEIVGKKNVLEBOPAGQDPSRLIHGCGISGERMDWAIQKTEIG VAETTPYSECEVRLKDERADKRLQHRQVAISACECCGSSPLIHPPALDMLR GRDDLRLVHPVAAPLASHAPPSRLAVLVGPEGLSDAEVEQAAGFHSARLGPV LRTEAPVVALVAQQLMGDA"
CDS	complement(2953..4356) /gene="b10a" /locus_tag="PA0420" complement(2953..4356) /gene="b10a" /locus_tag="PA0420" /note="Protein name confidence: Class 2 (high similarity

gene	to functionally studied protein); Alt protein name: 7,8-diamino-pelargonic acid aminotransferase DAPA aminotransferase"
CDS	/codon_start=1 /transl_table=11 /product="adenosylmethionine-8-amino-7-oxononanoate aminotransferase" /protein_id="AAG03809.1" /db_xref="GI:9946276"
gene	/translation="MGLNADMQRDLNVLMHPCTOKMDHRLPIPIRGEVWLEDF EGGRYIDAVSSWVWVFGHANPRINIKQVNDOLSHVILAGSHQVETLSRIYV TPGLDRVRYADSGACIEVALKMSYHFLNLSRPRPKRVITLNTSGHTIAAMSG DVALFTETYSLLDITKVPSDPCFLRPDSCWSEBSRKNFAMHRTLAGHDEIAV IVEPLIQAGAGMRYHPVYLKLREACDRGVHLIDELAVGRTGTMPACBOAGIA PDFLCISKALTGCVLPMSAVLTSETTYRGFYDDYOTLRAFLHSHYTGNIACAAALA TLIFIEEDRYTEANRALSTMARATNLADHPHVAEYRGTGMVLAIEVODKASRTY PMOERGLKVYFCHGLRGALRLRGLSVVYLPVYITPEQIDFLAEVASSGIDIAITRD AVSVAISDRPHPDHDPG"
CDS	4530..6020 /locus_tag="PA0421" 4530..6020 /locus_tag="PA0421" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAG03810.1" /db_xref="GI:9946277"
gene	/translation="MSLRMLRAAFLMLGVFSAVALGKOKOPTAIVGAGLAGISNAY ELGKDQVIVLEARPQVGGSGIATSEWGNQKQVPTLNAVYDTEFLKVPAPDFVR TLIDGLIGYSSDLAKOPNVAALDKRGESTLIDLSAISIDPLNPASSNTPLALDO MNAWMLDKNLSPTARILVNORIRSGYDEPSLSILYLAQOARVAGVDRDLRAAR LPGSGVLAEBAYKQIKITIKTSKVSIVQAKGVAVKASSETYKADYVYLAIPALAL GQIQTPPSLSGTQMSALKGTNYGWRDQILKFKRPVWDKSRISGETFSQDGMVW EPALKGANVILNSGDNARVILQAFDRQWVDVLLRMKFFYRMKGAFAGYAIRYS ADPGTGSYLAYPGQVTFEWMRIWEOPLSGVAGCHTDLAYGTTLEGALRSKRAAS QVRDLYAGKTPTVEGAPVMAKAEKVAFTTGCGSKNPFAMFTFLFD"
CDS	6201..6770 /locus_tag="PA0422" 6201..6770 /locus_tag="PA0422" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAG03811.1" /db_xref="GI:9946278"
gene	/translation="MOWRNTSRGYVFSPLFHWGSAVVGFLGLMLMBELSYDDPW YHPAPALHKSIGILLALILVRIYWRVSPPPRANHGGLTVASLGLALYGLL AVIVAGTLISIAUGEPISVFGMTSPVTTSLGPDQADVAGEIHLIYLMALVRAVLHA LAALKHIFVRDPTLKRMLGRSSNDQOPR"
CDS	6782..7357 /locus_tag="PA0423" /note="synonym: yceI" 6782..7357 /locus_tag="PA0423" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAG03812.1" /db_xref="GI:9946279"

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

3.24e-35 Length: 13987
753.00 Matches: 174
58.60% Conservative: 68


```
gene
/feature="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/feature="LT2"
70..900
/feature="cof"
/feature="synonym: STM0457"
70..75
/feature="cof"
/feature="putative RBS for cof; RegulonDB:STMS1H000802"
82..900
/feature="cof"
/feature="similar to E. coli orf, hypothetical protein (AAC73549.1); Blastp hit to AAC73549.1 (276 aa), 81% identity in aa 5 - 276"
/codon_start=1
/transl_table=1
/product="putative hydrolase"
/protein_id="AAL19412.1"
/db_xref="GI:16418965"
/translation="MARLAPDMGTLLMPDHHGRETATLARESDITLTPATGR
HYLEMRHLGTLSDAYLITGNGTRHSLBEDVLRHODLDQVADTVMHNAWDRASH
HVEDNNGMFTQEI PALIQAHVYSGFRYQVIDIKSIPAHQVTKLCPGSDHDLRLRI
QUNALEBRALICSAVDCLEVLPGCNKSGALAVLSNHLGLSLADCAFGDANNDRE
MLGSGVGRGLWGNAMPOLIALPHLSVIGHGNGQAVSHFLTHWLDNPLPYSP"
complement(950..2010)
/feature="STM0458"
complement(950..2005)
/feature="STM0458"
/feature="similar to E. coli cysteine synthase B, O-acetylserine sulfhydrylase B (AAC75474.1); Blastp hit to AAC75474.1 (303 aa), 26% identity in aa 6 - 207, 35% identity in aa 198 - 289"
/codon_start=1
/transl_table=1
/product="putative cysteine synthase/cystathionine beta-synthase"
/protein_id="AAL19413.1"
/db_xref="GI:16418966"
/translation="FMSSNMYKAININEINAHQGSADTHLRPLSAFPGIQLYKOE
STPTGLSKRLARSLFLYGLCNMIREGTPIDASSGSTAISAMFPRLLGLFLIAY
MPACTARKKIEIOIYGGCHHVSACEIYVAASERLAEHLNGHWDQPTVAREATDR
GNNIADSIPIROMENLHPVPIYMSAGTGTATGRIYRCQGYDTQMLVWPEVS
VFLPYMDRBRSLSPVSGKIEGIRPVEPSPIIDVVDMLRVPDASVATAMHLET
OLGRKVASCTGTMGALQILARMRREAGETAIYTLCDSDRLDITYYHRAWVSDHI
GDLTPWSALIKKLTGD"
complement(2005..2010)
/feature="STM0458"
/feature="putative RBS for STM0458; RegulonDB:STMS1H000803"
2107..2576
/feature="ybaO"
/feature="synonym: STM0459"
2107..2112
/feature="ybaO"
/feature="putative RBS for ybaO; RegulonDB:STMS1H000804"
2118..2576
/feature="ybaO"
/feature="similar to E. coli putative LRP-like transcriptional regulator (AAC73550.1); Blastp hit to AAC73550.1 (181 aa), 95% identity in aa 30 - 181"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator (asnc family)"
/protein_id="AAL19414.1"
/db_xref="GI:16418967"
/translation="MLDKIDRLCLLOQDCTLQALADAVNLTTPFCWRLRLLED
DGILLGRVALLDEPKLGLTAFVLIKTHGSSSMWCYFVAVSEMEPVIGFMMAE
VDYLMRYVADMKRYDDPYKRLVNSVPLSVPTSSFAMEQIKITSLPIE"
2604..4389
/feature="mdIA"
/feature="synonym: STM0460"
RBS
2604..2609
/feature="mdIA"
/feature="putative RBS for mdIA; RegulonDB:STMS1H000805"
2617..4389
/feature="mdIA"
/feature="similar to E. coli ATP-binding component of a transport system (AAC73551.1); Blastp hit to AAC73551.1 (590 aa), 88% identity in aa 1 - 590"
/codon_start=1
/transl_table=1
/product="putative ABC superfamily (atp) transporter"
/protein_id="AAL19415.1"
/db_xref="GI:16418968"
/translation="NRLFAOLSWYFRERREYGAVALMLMLQILPEKVGIVVD
GVTAOQPTPRIMMIGTALIAVVYLLRYVRLFGASVQALVARELDYRQSLR
OHPEFYORHRTGDLIARATNDVDRVFAAGEGLTVDSLVMGCAVLWSTQISML
TLIALPMPIMALMIKRYGDRLDYFKLQAARSSANDRQESTSTIRMKAGLEDR
OSSLFADADDTGKNNRVARIDARFDPTIYIAGNALIASGSMVNVGSLTIGE
LTFPMYVIGIMIPMLALAMFNI VERGSAAYSRIPAMLEAVVDGEPVPAQGE
LTPAIRFCYPTHTHPALEVNFRLPGOMLIGCPGAGKSTLSLQHPVTOGE
IRPHMDPLHLODSWRSRLAAYVORTPEFSDSIANNIALGREATOEBIEQVRLAS
VHEDILRLPGCVTOVGERGMVSGGOKORISIALPALLNAITLLDLSAIVAGRE
HOLHNIRQNGERTVIYISNHRSLATDANEIITVMQGHVYQGDHQAOLQGVARD
MYRYQLEALDPAERGEBAVNA"
4369..6163
/feature="mdIB"
/feature="synonym: STM0461"
4369..4374
/feature="mdIA"
/feature="putative RBS for mdIB; RegulonDB:STMS1H000806"
4382..6163
/feature="mdIB"
/feature="similar to E. coli putative ATP-binding component of a transport system (AAC73552.1); Blastp hit to AAC73552.1 (593 aa), 88% identity in aa 1 - 593"
/codon_start=1
/transl_table=1
/product="putative ABC superfamily (atp membrane) transporter"
/protein_id="AAL19416.1"
/db_xref="GI:16418969"
/translation="MRSFGQMLPTLRLALVYSGPWKPLSAVMMIIMIAAEVSGPL
DLVYFDLNMVARNHPLPGKVAALAAVYGLQFLAAGLHAQSLIFRPAAGVQSLST
DVAADLROGLSFLDPOVQGLISRTVDNTEYRDLYVAVLSALIGMLVAM
FISLQDMRALVALIIPAVLITVMIYORSTPIYRRAVYADINDGENIINGSVTO
OPROQARCGERMGESASRSHYMAVOTLRDGLLRPLSLFSALLICGLMLPSFSA
GITTEGVLYAFIYLSRLNPLELTTOQSMQOAVYAGREVELNDPRQYSGSDR
PLQSGAIDIDHLSFAYRDDNLVQDITLSVPSRSFVALVGHTSKSTLSLNGYTP
LTQGEIRLDGREIASLNHRYLRQGVAMVQDDPVVMDTFLANVTLGRDSEAOVQWL
ETVQLADLAGLSDGLHTLHGEQNTLSVGKOLLALARYLVADAPQILIDETASID
SGTEOAIQCALAIRERTTLVIAHRLSTIVEADTIVLRGQAVERTGTHQQLAAG
RYWQMOVQOLVGBELAASVHEBGPAA"
6363..6713
/feature="glnK"
/feature="synonym: STM0462"
6363..6368
/feature="glnK"
/feature="putative RBS for glnK; RegulonDB:STMS1H000807"
6375..6713
/feature="glnK"
/feature="nitrogen assimilation by glutamine synthetase; regulates Gln (NRII) and GlnF (Glnase); similar to E. coli nitrogen regulatory protein P-II 2 (AAC73553.1); Blastp hit to AAC73553.1 (112 aa), 97% identity in aa 1 - 112"
/codon_start=1
/transl_table=1
/product="regulatory protein P-II 2"
/protein_id="AAL19417.1"
/db_xref="GI:16418970"
/translation="MKLYVYIIRKPLFEDVREALSGTIGLTVTVEYKGRGKHAE
LYRGAETSVNPLPRKVIDAIVDDQLEVIDITISKAAYTKIGDGKIFVAILRWVIRI
RTGEADEAAL"
6731..8031
gene
```

US-09-889-756A-2 (1-412) X AE008717 (1-20347)

Db	18554	ACGGCAATTACGTTCCACAACAGCGCGTTAACCCGTACTCACAAGCGGCATGCCACGGTG	18493
Oy	329	MettIleValAsmAlaGlnGlyMetSerGluProArgIuValThyValAlaGlnGlnGln	348
Db	18494	CTGGTGGTGGCCCTCATATAACAAGGAAACCCTGCAATGCTGCGAAGCCAGCGCATC	184359
Oy	349	G1ThrAsnTrpLLeValThrSerGlyLeuLysAspGlyAspLysValValGluGly	368
Db	18434	GCGGATTAAGTGGCTGTGACTGACGATTCGAAAGCGGCGCACCGCTAGTCTGTACGGGG	183757
Oy	369	IleSerIleAlaGlyIleThrGlyAlaLysLysValThrPro-----	382
Db	18374	CTG-----CAAAGAATACGTCCTGGGCGACAGGTTAAAGTA	183339
Oy	383	-----LysGIUTrpAlaSerSerGluAsnGlnAlaAlaPro	395
Db	18338	CAGAAATTACCGCGGATTAACAACAAGCAGCAGCGGATCAATCCTCAGCCCC	182797
Oy	396	GlnSer	397
Db	18278	AGGTCT	18273
RESULT 31			
ECOACREF		4571 bp	DNA linear BCT 13-OCT-1993
LOCUS			
DEFINITION	E. coli acriflavine resistant proteins (acrB and acrF) genes.		
ACCESSION	complete cds.		
VERSION	M96848		
KEYWORDS	acrf gene; acrf gene; acriflavine resistance protein; cytoplasmic membrane protein.		
SOURCE	Escherichia coli		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (sites)		
AUTHORS	Klein,U., Henrich,B. and Plapp,R.		
TITLE	Molecular cloning of the envC gene of Escherichia coli		
JOURNAL	Curr. Microbiol. 21, 341-347 (1990)		
REFERENCE	2 (bases 1 to 4571)		
AUTHORS	Xu,Y. and Bertrand,K.P.		
TITLE	Nucleotide sequence of the acrf operon from Escherichia coli		
JOURNAL	Unpublished (1992)		
COMMENT	On Oct 13, 1993 this sequence version replaced gi:290407. Original source text: Escherichia coli (strain K-12) DNA. Location/Qualifiers 1..4571 /organism="Escherichia coli" /strain="K-12" /db_xref="taxon:562" 191..194 /gene="acrB" 202..1359 /gene="acrF" 202..1359 /gene="acrF" /function="resistance to acriflavine, hydrophobic antibiotics, basic dyes, and detergents; suppresses acrA and acrB mutations" /codon_start=1 /transl_table=11 /product="acriflavine resistance protein" /protein_id="AAA02931.1" /db_xref="gi:290408"		
FEATURES			
source			
RBS			
gene			
CDS			
sig_peptide			


```

/note="reference_sequence = X57948"
/citation={1}
/replac="at"
BASE COUNT      1105 a      1114 c      1209 g      1143 t
ORIGIN
Alignment Scores:
Pred. No.:      1,06e-35      Length:      4571
Score:          750.50      Matches:      168
Percent Similarity: 60.65%      Conservative: 74
Best Local Similarity: 42.11%      Mismatches: 122
Query Match:    37.17%      Indels:      35
Db:             1      Gaps:        6
US-09-889-756a-2 (1-412) x ECOACREFF (1-4571)
QY      17 AlaValAlaLeuValLeuSerSerCysGlyGlyGlyAspAlaAlaGlnGlyGln 36
      250 TCCCGGCGTTTAAATGCGCGTTTAAAGGAGAGAAAGCTCACGTCGTTAA 309
QY      37 ProAlaGlyArgGluAlaProAlaProValAlaGlyValThrValHis---ProGln 55
      310 CCGGAG-----GTTACCGTTCAATTTGTAATAA 336
QY      56 ThrValAlaLeuThrVal-----GluLeuProGlyArgLeuGluSerLeuArgThrAla 73
      337 ACGGCGCGTTAAGAAAGTAAGAGTAATTAACAGGCGGCAACCAATGCTTATGTAAGCC 396
QY      74 AspValAlaArgAlaGlnValGlyGlyAlaGlnGlnValGlnValGlnValGlnValGlnVal 93
      397 GAAGTTCGCCCAACGAGTTAAGGAGATCTGATCGCAATTCACATGAAAGGACGAT 456
QY      94 ValArgAlaGlyGlnProLeuThrGlnLeuAspSerSerThrTrpGluAlaAsnLeuGlu 113
      457 GTGCAAGAGGCGCGCTCTTACCAATTCATCCCGGACCTTACGCGCAATTAATGAC 516
QY      114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaValAlaAspAlaAspLeu 133
      517 AGCGCGAAAGCGAAGCTGCGAAGAAAGCGCGCGCGCATCGCGCATTTGACGTA 576
QY      134 AlaArgTrpTrpProLeuValAlaAlaGluAlaValSerArgGlnGlyTrpAspAlaAla 153
      577 AAAGGTTACGTTCCGCTCGTGGTACGAAATACATCCGCGAGGAGTACGACGACG 636
QY      154 ValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaLeuSer 173
      637 ATTGCTGATGCTCTTACAGCGCGAGTGCCTGATTCGCGCAACGACGTAAGAGC 696
QY      174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln 193
      697 GCTCGCATCAATCTTGGCTTATACCAAAAGTCACTCGCGCAATTAACGAGCGTATCGGCAA 756
QY      194 SerLysValSerGlyGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213
      757 TCGACTGTGACCGAAGGCGCTTGTCTCATATAGGCAACGACTGAACTGGCGACGTC 816
QY      214 ArgGlnThrAsnProMetTrpValAlaAsnValThrGlnSerLysArgGlnValMetLysLeu 233
      817 CAGAGCTGATCTTATCTTATGATGATGACCCCAATTCAGCAACGATTTATAGGCTG 876
QY      234 ArgArgGlnIleAlaGlyGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIle 253
      877 AAGCAATCCGTAAGCAAGAAATTTGATTAAGAAAGCCACCAACGATGAGATTG 936
QY      254 LysPheAspArgGlyThrValTrpProGlyLysGlyArgLeuLeuPheAlaAspProVal 273
      937 GTCATGAAAAACGCTCAAACTATCCCTGAAAGTACGCTGCAATTCCTCGATGTGACC 996
QY      274 ValAsnGlySerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
      997 GTTATAAAGACACCGGCTCATTAACCTTACGTGCTCTTCTTCCCTAACCAGCAACGACG 1056
QY      294 LeuMetProGlyLeuTrpValArgValIleuMetAspGlnValAlaValAspAsnAlaPhe 313

```

```

Db      1057 CTTTTCGGGATGATTTGGTCGTCACGAGTATGATAGGCGTCAACCTGACGCATT 1116
QY      314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
      1117 CTTATCCCGCAACAAAGCGGTAGCGCACACCGCGGTGATGACACCGGTGCTAATGTT 1176
QY      332 AsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsn 351
      1177 AACCATAAAGTCAAGTTGAAGCCCGCTGCTTCCAGTCAAGCGATTTGGCGATTA 1236
QY      352 TrpIleValThrSerGlyLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIle 371
      1237 TGGTTGATTAGTGAAGCATGAATCTGCGGATCAAGTCAATTCACAGCGGCTG----- 1290
QY      372 AlaGlyIleThrGlyAlaLysLysValThrProLysIleThrAlaSerSerGluAsnGln 391
      1291 -----CAAAAACCGCTCCGGAGAG----- 1311
QY      392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
      1312 -----CAGTTAAGCACATCCGATACCCCGCAGATCTGATGTAAG 1356
Db
RESULT 32
ECENVC D      6803 bp      DNA      linear      BCT 07-JUL-2002
LOCUS      E.coli envC, envD and envr genes.
DEFINITION      X57948
ACCESSION      X57948
VERSION      X57948.1 GI:510827
KEYWORDS      envC gene; envD gene; envelope protein; envr gene; insertion
              element; insertion element 2.
SOURCE
ORGANISM      Escherichia coli
              Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS      Klein,J., Heinrich,B. and Plapp,R.
TITLE      Molecular cloning of the envC gene of Escherichia coli
JOURNAL      Curr. Microbiol. 21, 341-347 (1990)
REFERENCE
AUTHORS      Klein,J.R., Heinrich,B. and Plapp,R.
TITLE      Molecular analysis and nucleotide sequence of the envCD operon of
              Escherichia coli
JOURNAL      Mol. Gen. Genet. 230 (1-2), 230-240 (1991)
MEDLINE      92079901
PUBMED      1720861
REFERENCE
AUTHORS      Seifert,D., Klein,J.R. and Plapp,R.
TITLE      EnvC, a new lipoprotein of the cytoplasmic membrane of Escherichia
              coli
JOURNAL      FEMS Microbiol. Lett. 107 (2-3), 175-178 (1993)
MEDLINE      93231461
PUBMED      8472900
REFERENCE
AUTHORS      Klein,J.
TITLE      Direct Submission
JOURNAL      Submitted (22-FEB-1991) Klein J., Universitaet Kaiserslautern
              Erwin-Schroedingerstr. 23 Kaiserslautern, Germany
REMARK
AUTHORS      revised by [7] MAT
TITLE      5 (bases 1 to 6803)
JOURNAL      Direct Submission
COMMENT      Submitted (12-JUN-1994) Klein J., Universitaet Kaiserslautern
              Erwin-Schroedingerstr. 23 Kaiserslautern, Germany
              On Jul 17, 1994 this sequence version replaced gi:41350.
FEATURES
source
1. 6803
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/complement(42..704)
/note="in vitro and in vivo expression of cloned envr

```

CDS

```

gene
/codon_start=1
/transl_table=11
/product="potential repressor envr"
/protein_id="CAA1015.1"
/db_xref="GI:510828"
/db_xref="SWISS-PROT:P31676"
/translation="MAKRTKALKTQELLETALIAOPAGHGSKTLINDIADANVT
RGAIVHPEKNTOLFNEMWLOPSELRLIOHLEHDPFOOLREKTLVIGQYIAK
IPROQALIKTYHKCEPNDKMLAGVAREKGRPNPOTIRVLOACCOQGCYANNLDD
VMTIIDAFSGIVQNMMLNMAGYDLKQAPALVDNVMRMPDENITKLIHONTELS
VM"
repeat_region
1018..2344
/insertion_seq="Insertion element 2, IS2(II)"
promoter
2321..2353
/citation=[4]
/citation=[3]
/evidence=experimental
mRNA
2356..6803
/citation=[4]
/function="envCD messenger RNA"
join(2425..2428,2435..3592)
/gene="envC"
2425..2428
/gene="envC"
2428..2428
/citation=[4]
/citation=[3]
2435..3592
/gene="envC"
/citation=[4]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="envelope protein"
/protein_id="CAA1016.1"
/db_xref="GI:510829"
/db_xref="SWISS-PROT:P24180"
/translation="MTKARPELLPSFLISALIAAGCNDKEERKAHVPOVTHIV
KTAFLVKTLEPGRITNAVRIAEVRPOUSGILNMFTEGSDVOAGOSLVOIDPTVIA
NYDSKGEIASEPAAAIHHTVTRVYPLVETKTSIOEYTOAIADARQADANVIAK
ATVESARINLAVITVAVTISGRICKSTVBGALVNGOTELATVOQIDPIYDVTS
SNDPMRLKOSVEQGNLHKNATSNVELVMENGQYTPKGTLOFSDVDESTGILTR
AVFPNPOTHLPLGMFVARIDEGVOPAILIPQOGVSRTPGDATTVLIVNKSQVEAR
PVVASQALIGDKMLISEGLKSGDYIVISGLQKARPEQVATTDPADTASK"
2435..2503
/gene="envC"
/citation=[4]
2504..3589
/gene="envC"
/function="envelope protein, envC"
/function="cytoplasmic membrane protein"
/note="mature protein is found exclusively in the
cytoplasmic membrane"
/citation=[4]
/evidence=experimental
join(3592..3594,3604..6708)
/gene="envD"
3592..3594
/gene="envD"
/citation=[4]
/citation=[3]
3604..6708
/gene="envD"
/function="cell division"
/citation=[4]
/codon_start=1
/transl_table=11
/protein_id="CAA41017.1"
/db_xref="GI:510830"
/db_xref="SWISS-PROT:P24181"
/translation="MANFTRPIFAWLAIIIMAGALAILQIPVAQPYTIAPPAVS
AVSAPGADQTVODTVQVLEONNMNIDIMWSSTDSAGSVTILITLPOSGTDPPI
AOVONKTLATPLPQEVQOGISVEKSSSYIMVAGFVSDNPGTTODDISDVAS

```

```

NRDQTLRLNGVDVOLFQAQVAMRIWLDADLNTKYLPVDVINOLKQNDQIAAQ
LGGTALPAGQOLNASTIIACTRFKNPEEFGKVTLRVNSDGSVRLKQVARELGSENN
VIARINKPAPAGIGIKIATGANNALDIAKAKIATLAEVDSGVSRLKQVARELGSENN
LSIHEVKTITFEAIMVFLVMVFLFIONMRATLPTAVPVLGTLVLAIRGYSINT
LTMFWALIGLAVDDAIVVENVERWMMEDSKPPREKESQIOGALVGLAMTIS
AVPIPMAPFGSGTGAIVROPSITIVSAMALSVAILLPALCATLLKPVASHEK
GGFFGFMNTTFDHSVNHVNSVGKILGSTRYLLIYALVAGVVFLEPLPSSELPPE
DOGVFPLMIDLPAGATOEKQVLDQVTLKXENKAVSEVFTVNSFGQAQNGAG
MAVSVLKPMEERNDENSAEVHRAKMEGLIRDPVLPENNPATVEIGETAPDE
LIDQAGIGHALPOARNOLGMAOCHPASLVSVPNGLEPTAQPKLEVDKQALGV
LSLDINDITISTALGTYVNDFTIDGRVKKLYVADAKFRMLPBDVCLKYRSANGEV
PFSAFSTSHWVSGPRLERYNGLPSEMIQGEAPGTSGDAMALMENTALPAGVIA
DWTGMSYOEBSHNSQSPALVAISFVVVFLCALVYSWSIPVSMVLPVIGVLLA
ATLPKQNDVYFMVGLTLTGISAKNALIIVEPAKLMKEKGVVATIMAYMRIR
PIIMTSLAFLIGVLPLAISNGAGSQANNAIGVGMGVSATLLAIFVVFVFIHR
CFKQ"
sig_peptide
3604..3681
/gene="envD"
/citation=[4]
BASE COUNT 1706 a 1612 c 1764 g 1721 t
ORIGIN
Alignment Scores:
Pred. No.: 1,78e-35 Length: 6803
Score: 750.50 Matches: 168
Percent Similarity: 60.65% Conservative: 74
Best Local Similarity: 42.11% Mismatches: 122
Query Match: 37.17% Indels: 35
DB: 1 Gaps: 6
US-09-889-756a-2 (1-412) x ECENVC (1-6803)
Qy 17 AlavAlaLeuValLeuSerSerCysGlyYleGlyValAspAlaIaGlnGlyGln 36
Db 2483 TCCGCGGCTTTAATCCCGCGGTGTAACGATPAGGAGAAAGAGCTCACGTCGGTGA 2542
Qy 37 ProAlaGlyArgGluAlaProAlaAlaValGlyValThraValHis--ProGln 55
Db 2543 CCGGAG-----GTTACCGTTCATATTGTAA 2569
Qy 56 ThrValAlaLeuThrVal-----GluLeuProGlyArgLeuGlnSerLeuAlaGln 73
Db 2570 ACGGCCCCGTGAGAGTTAAGACTGAATTAACAGCCGACCAATGCTTATCGTATGACC 2629
Qy 74 AspValArgAlaGlnValGlyGlyIleGlnLysArgLeuPheGlnGlySerTyr 93
Db 2630 GAAGTTGCCCCACAGGTTAGCGGATGATCTGAATGCCAATTTCACTGAAGCCAGCAT 2689
Qy 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGlnAlaAsnLeuGlu 113
Db 2690 GTCCAGACAGGCGAGTCCCTGTACCATCGATCCCGACCTATCAGGCAAAATTATGAC 2749
Qy 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeu 133
Db 2750 AGCCCGAAGCGCAACGCGGAAAGGAGCCGCCGCCCATCGGCAATTGACCGGTA 2809
Qy 134 AlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGlnTyrAspAlaAla 153
Db 2810 AAACGTTACGTTCCGCTCGTGCGTACGAAATACATCCAGCAGCGAGTACGACAGCC 2869
Qy 154 ValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173
Db 2870 ATTGCTGATGCTGTCAGCGCATGCGCGCGTATTCGCCAAAGCCACAGTCGAAAGC 2929
Qy 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGlyGln 193
Db 2930 GCTCGCATCATCTTCTTATACCAAGTACATCGCCGCAATTGCGGACGATCGGCAAA 2989
Qy 194 SerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213
Db 2990 TCGACTGTGCGAAGCGCTCTTGCTACTTAAGGCAACGATGACGACGACGCTGC 3049
Qy 214 ArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeu 233

```

```

Db      3050 CAGAGCTTCATCTTACGTGATGATGACCAATCCAGCAACGATTTATAGCGCTG 3109
Qy      234 ArgArgGlnIleAglUGlyLeuLeuLeuAlaIleAspGlyValIleAlaValGlyIle 253
Db      3110 AAGCAATCCGAGGCAAGAAATTTGATAGAAAAGCCAGCAGCAGTGAAGTTG 3169
Qy      254 LysPheAspAspGlyThrValTyProGlnIleGlyArgLeuLeuPheAlaAspProVal 273
Db      3170 GTCATGAAAACGGTCAACCTATCCCTGAAAGTACCGTCAATTCCTCGATGTACC 3229
Qy      274 ValAsnGlnSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db      3330 GTTATGAAAGCACCAGCTCCATACCTACGTCTGCTTCTTCCCTAACCCGCAACATACG 3289
Qy      294 LeuMetProGlyLeuTyValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db      3390 CTTTGGCCGGGTATGTTTGTCGTGACAGGATGATAGAGCGTCCAACTGACGCCATT 3349
Qy      314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
Db      3350 CTTATCCCGCAACAAGCGTTAGCCGACACCGCGGTGATGACACCGCTGATTGTT 3409
Qy      332 AsnAlaGlnGlyMetGluProArgGluValThrValAlaGlnGlnGlyThrAsn 351
Db      3410 AACGATAAAGTCAGTTGAAGCGCCGCTGCTGTCAGTACGAGCGATGGCGATAAA 3469
Qy      352 ThrIleValThrSerGlyLeuLysAspGlyAspLysValValValGlyGlyIleSerIle 371
Db      3470 TGGTATGTTGTTGAAGACGTAATCTGGCGATCAAGTCAATGTCACCGGCGCTG 3523
Qy      372 AlaGlyIleThrGlyAlaLysLysValThrProLysGluThrPalaSerSerGluAsnGln 391
Db      3524 -----CAAAAAGCGCGTCCGGAAG----- 3544
Qy      392 AlaAlaIleProGlnSerGlyValGlnThrAlaSerGlyAlaLysThrAlaSerGlu 410
Db      3545 -----CAGTTAAAGCCACTACCGATACCCCGAGACTGCTGATCGAAG 3589

RESULT 33
LOCUS      AE000405              11095 bp      DNA              linear      BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 295 of 400 of the complete genome.
ACCESSION  AE000405.1  GI:1789659
VERSION     AE000405.1  GI:1789659
KEYWORDS
SOURCE
ORGANISM    Escherichia coli K12
             Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
             Enterobacteriaceae; Escherichia.
REFERENCE   1 (bases 1 to 11095)
AUTHORS    Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Blythe,M., Collins,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
TITLE       The complete genome sequence of Escherichia coli K-12
JOURNAL     MEDLINE
PUBMED     97426617
REFERENCE   2 (bases 1 to 11095)
AUTHORS    Blattner,F.R.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3 (bases 1 to 11095)
AUTHORS    Blattner,F.R.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:

```

```

608-263-7459
4 (bases 1 to 11095)
Plunkett,G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
         source
         1..11095
            /organism="Escherichia coli K12"
            /mol_type="genomic DNA"
            /strain="K12"
            /sub_strain="MG1655"
            /db_xref="taxon:83333"
         7..35
            /note="factor sigma70; promoter fls; documented +1
            at3407883"
         76..1041
            /gene="yhdG"
            /note="synonym: b3260"
         76..1041
            /gene="yhdG"
            /function="putative enzyme; Not classified"
            /note="o321, 100 pct identical amino acid sequence and
            equal length to YHDG_ECOLI SM: P25717"
            /codon_start=1
            /transl_table=11
            /product="putative dehydrogenase"
            /protein_id="AAC76292.1"
            /db_xref="GI:1789660"
            /translation="NRIGYQLRNRIAPMAGITRPPRTLCYEGAGITSEMMSS
            NPGWESDSRLRMVHIDPGRIVQIASDPEEMADARINVESAOIIDIMGCPA
            KYNRKLASALQYDPDVVKSITLVEVNAVDPVTLKIRGMAPBEHRNCEIADYLD
            CGIOLITIGRTTRACIFNGEAEYDSIRAVKVSIPVINGDITDPLKARAVADYGA
            DALMTGRAAGPWIIFREIOHYLDGELLPLPIAVKRLCAHVEILDHPGPAGY
            RIARKHVSWYLOEHANPDQFRRTFNAINEDASQLELAEVFNFA"
         946..972
            /gene="yhdG"
            /note="factor sigma70; predicted +1 start at 3408820"
         1067..1363
            /gene="fls"
            /note="synonym: b3261"
         1067..1363
            /gene="fls"
            /function="factor; DNA - replication, repair,
            restriction/modification"
            /note="o98; CG Site No. 18328"

```


ACCESSION AE005554 AE005174
 VERSION AE005554.1 GI:12517891
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 1 (bases 1 to 11807)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,Y., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)
 21074935
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 11206551
 2 (bases 1 to 11807)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,Y., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES
 SOURCE
 1. 11807
 /organism="Escherichia coli O157:H7 EDL933"
 /mol_type="genomic DNA"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 70. .960
 /gene="yhjJ"
 /note="synonym: Z4622"
 70. .960
 /gene="yhdJ"
 /function="putative enzyme; Not classified"
 /note="Residues 1 to 296 of 296 are 98.98 pct identical to
 residues 1 to 296 of 296 from Escherichia coli K-12 Strain
 MG1655: B3262"
 /codon_start=1
 /transl_table=11
 /product="putative methyltransferase"
 /protein_id="AAG58390.1"
 /db_xref="GI:12517892"
 /translation="MTMRGCEPTPRFGNEAKTIIHGDALAEKKLPESVDLIPADPP
 YNIGKMPGLIEAWKEDLFIDMLLEVAIEGHRVLEKQSGWYINSTEMPTIDLOCKR
 LFTIKSRIVMSDSSGVQAKGYGSMERPIIMVKADAKNTYFNDALIVEKTSQQA
 LIDRKNPQPNYHOKVGNWDFPRVRLYDEENHTQKPEALKKIIILASSNPGD
 IVLDPFASFTGAVAIASGRKFIGIEINSEYIKWGLRRLLDVASHYSABELAKVKKSK
 TGNLSKSRSLSEVPDLITK"
 1044. .1223
 /gene="yhdJ"
 /note="synonym: Z4623"
 1044. .1223
 /gene="yhdJ"
 /function="orf; Unknown function"
 /note="Residues 1 to 59 of 59 are 100.00 pct identical to
 residues 1 to 59 of 59 from Escherichia coli K-12 Strain
 MG1655: B3263"
 /codon_start=1
 /product="orf, hypothetical protein"
 /protein_id="AAG58391.1"
 /db_xref="GI:12517893"
 /translation="WIRKYMWLVVPAVAVFLPFTLLMQWIELLATETDKCKNNVNP
 LKLVNCELENFQDRM"
 complement(1226. .1888)

CDS
 /gene="envr"
 /note="synonym: Z4624"
 complement(1226. .1888)
 /gene="envr"
 /function="putative regulator; Not classified"
 /note="Residues 1 to 220 of 220 are 100.00 pct identical
 to residues 1 to 220 of 220 from Escherichia coli K-12
 strain MG1655: B3264"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator"
 /protein_id="AAG58382.1"
 /db_xref="GI:12517894"
 /translation="MAKRTAEELKTRQELIETAIAGPAQGVSKTTLNDIADAANT
 RGAIVHFEKTKTQLEFEMMLQPSLELIEGHTAGIEHPFOOLREKLIYGLQYIAK
 IPROALKLIIYHKCEPNDMLAEGVIREKMGVNPOTLRVLAOCQCGCVANNLDD
 VMIILIDAFSGIVQNMVLMNAGVDLYKQAPALVDNVLRFMFDENTIKLIHQINELS
 VM"
 2287. .3444
 /gene="acrE"
 /note="synonym: Z4625"
 2287. .3444
 /gene="acrE"
 /function="membrane; Cell envelop: Inner membrane"
 /note="Residues 1 to 385 of 385 are 99.74 pct identical to
 residues 1 to 385 of 385 from Escherichia coli K-12 Strain
 MG1655: B3265"
 /codon_start=1
 /transl_table=11
 /product="transmembrane protein affects septum formation
 and cell membrane permeability"
 /protein_id="AAG58393.1"
 /db_xref="GI:12517895"
 /translation="MTKHARFLLPSRLISALLIAGCNDKGKHAHVGEPOYVHV
 KTRPLEKTELPGRNTAAYRIAEVRPOVSGIVLARNPREGSDVQAGSLYVIDPATYA
 NYSAKGEILAKSSAAAIATLVKRYVPLGTYISQOETDQIADARQDAVATLAK
 ATVESARINLAYTKVTAIPISGRIGKIVTEGALVINGQTTETLAIVQOLDPIYDVYTS
 SNDPMRLKQSEQGNLKHENATNVELVMENGGQTPYLGKTLQPSDVLVDESTSTLR
 AVFPNPQHTLLPGMFVARIDEGVDPALIPQGSRTFRGDAVLIIVDKSGOVEVR
 PVVASQAGIDKMLISGLKSGDQVIVSGIQAKARPGVAKTTPADTAK"
 3456. .5342
 /gene="acrF_1"
 /note="synonym: Z4626"
 3456. .5342
 /gene="acrF_1"
 /function="putative membrane; Protection responses:
 Drug/analog sensitivity"
 /note="Residues 1 to 620 of 628 are 99.19 pct identical to
 residues 1 to 620 of 1034 from Escherichia coli K-12
 strain MG1655: B3266"
 /codon_start=1
 /transl_table=11
 /product="N-terminal fragment of integral transmembrane
 protein involved with acridine resistance"
 /protein_id="AAG58394.1"
 /db_xref="GI:12517896"
 /translation="MANFIRRPDIPAVLAILMMAGALILQLPVAYPTIAPPAVS
 VSAHYRGADQGVDTVTVIVIRONNGINDLMSSSDSAGSVITLITROSGDDPI
 AQVQVQKQLATPLPQEVQOGISVEKSSSYLMVAGVSNPDTODDISDYAS
 NVKDTLSRLNSVDVQLFGAYAMRIWLDLNLKTLVTVDIVINOLIKVNAQAIAAQ
 LGGTALPGQOLNASIITACTRLNKEBFQVTIRVSDSGSVRLKQVARELAGEENV
 VIARINKPAAGIGIKLATGANALDTAKAKLAELQEPFGQMKVLVYDTPFVQ
 LSIHEVYKTLFEALIMVLVIMVFLQNMRAITLPTIAVPVLGTFALIAAFYSINT
 LTMFGWALIGLIVDAIVVENVERVMEDEKLPKREATKSKSQIGALVGLAMVLS
 AVFTPMAPFGSGTGAIRPQSTITVSMAISVALILVLPALCATILKPTPSAFLEK
 GFGFMFNTTFDSVNHVYNSVGKILGSTRLIILVAILVAGNVVFLRLPSSEFLPEE
 DCGVFLTMQLPAGATQERTQKVLVDVTVYLLKMERANVESVTVNGFSFGAPAPOP
 HVT"
 5317. .5369
 /note="O-island #129; Region of the EDL933 chromosome not
 homologous to E. coli K-12 MG1655"
 5373. .6605
 /gene="acrF_2"
 misc_feature
 gene

```

CDS
/feature="synonym: Z4627"
5373..6605
/feature="acrf 2"
/function="putative membrane; Protection responses:
Drug/analogue sensitivity"
/note="Residues 1 to 410 of 410 are 99.02 pct identical to
residues 625 to 1034 of 1034 from Escherichia coli K-12
Strain MG1655: B3266"
/codon_start=1
/translation=1
/product="C-terminal acridine of integral transmembrane
protein involved with acridine resistance"
/protein_id="AAG58395.1"
/db_xref="GI:12517897"
/translation="MAFVSLKPEWERNDENSAEAVIHRAMELGRIDGFVPMNP
AIVELGTAGDFELIDKAGIDALTOARNOLLGMAAHPALVSVRPGLDITOF
KLEVDOKAOLGSLSDINOTSTALGCTYVNDPDRGKRYVVOAARAPMLPED
VDLIVRSTNGENVPPSAPFTTSHWYVSPLEENGLPSHEIGELAPGTSGDNAL
MENLAKLPAGIGYDWTGMSYERLSGNAPALVAISFVVFCLALVESWSIPVS
MLVPLVIGIVLTLATLENQKNDVYEWLTTIGLSAKNALIVFADKLMKEGK
VVEATMAVAMRLRPILMTSLAFLGVLPLAISNGASGQNAVIGVGMVSAITLL
AIFPVXFEVIVIRCFKG"
6858..7079
/feature="synonym: Z4628"
/feature="yhdV"
/feature="yhdV"
/function="orf; Unknown function"
/note="Residues 1 to 73 of 73 are 100.00 pct identical to
residues 1 to 73 of 73 from Escherichia coli K-12 Strain
MG1655: B3267"
/codon_start=1
/translation=1
/product="orf, hypothetical protein"
/protein_id="AAG58396.1"
/db_xref="GI:12517898"
/translation="MKRLIPVALITLALLACAHSPCVPYDDOGLVHTNTCKGTT
QDNWETAGAIAGAGAAVAAGLTMGIIALSK"
7618..8535
/feature="yhdW"
/feature="yhdW"
/note="synonym: Z4629"
/feature="yhdW"
/function="putative transport; Not classified"
/note="Residues 1 to 305 of 305 are 99.67 pct identical to
residues 1 to 305 of 305 from Escherichia coli K-12 Strain
MG1655: B3268"
/codon_start=1
/translation=1
/product="putative periplasmic binding transport protein"
/protein_id="AAG58397.1"

Alignment Scores:
Pred. No.: 4.2e-35 Length: 11807
Score: 749.50 Matches: 168
Percent Similarity: 60.65% Conservative: 74
Best Local Similarity: 42.11% Mismatches: 122
Query Match: 37.12% Indels: 35
DB: 1 Gaps: 6

US-09-889-756a-2 (1-412) x AE005554 (1-11807)
QY 17 AlaValAlaLeuValLeuSerSerCysGlyLysGlyValAspAlaAlaGlnGlyGln 36
Db 2335 TCCGCGCGCTTAATCCGCGGTGTAACGATPAGGAGAGAAAGCTCACTCGGTGAA 2394
QY 37 ProAlaGlyArgGluAlaProAlaProValValGlyValThrValHis--ProGln 55
Db 2395 CCGGAG-----GTTACCGTTCAATTGTAATA 2421
QY 56 ThrValAlaLeuThrVal-----GluLeuProGlyArgLeuGluSerLeuAlaGlnAla 73
Db 2422 ACGGCGCGCTTAAGAACTGATTAACGAGCGCGCACCAATGCTTATCGTATAGCC 2481

```

```

QY 74 AspValArgAlaGlnValGlyGlyIleIleGlnLysArgLeuPheGlnGlySerTyr 93
Db 2482 GAAGTTCCGCCACAGGTTAGCGGATCGTACTGAATCGCAATTCTACTGAAGCAGCAT 2541
QY 94 ValArgAlaGlyGlnProLeuTyrGlnIleAspSerSerThrTyrGluAlaAsnLeuGlu 113
Db 2542 GTCCAGACAGGCGCAGTCCGTGTCAGATCGATCCCGCAGCTATACGGCAAAATTATGAC 2601
QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeu 133
Db 2602 AGCGCCAAAGGCAACTGGCGGAAAGTGAAGCCGCCGCCATTCGCGCATTTGACGGTA 2661
QY 134 AlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnLysTyrAspAlaVal 153
Db 2662 AAGCGTTAGCTTCGCTCGGGGTAGCAATATCATCAGCAGCAGAGTATGACAGGCC 2721
QY 154 ValThrAlaLysArgSerAlaGluAlaGlyValLysAlaAlaGlnAlaAlaIleLysSer 173
Db 2722 ATTCGTGATGCTTCGCGAGCGCATGCGCGGTGATTCGCGCAAAACGACGTGAAAGC 2781
QY 174 AlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSerGlyPheIleGln 193
Db 2782 GCTCGATCATCTTGTCTTATACCAAGTACTGCCCAATTAGCGGACGTATCGCAAA 2841
QY 194 SerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThrValLeuAlaThrIle 213
Db 2842 TCGACGTGACCGCAAGCGCTCTTGCATCAATAGGCGCAACGACTAAGCGCAACTGTC 2901
QY 214 ArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeu 233
Db 2902 CACGACCTGCATCTTATCTTATGTTGATGATGCCAACCAATCCGCAACGATTTTATGAGCTG 2961
QY 234 ArgArgGlnIleAlaGlnGlyLysLeuLeuAlaAlaAspGlyValIleAlaValGlyIle 253
Db 2962 AAGCACTCGTATAGAGAGAAATTTGCATPAGAGAAACCCACGACGAACTGAGACTTG 3021
QY 254 LysPheAspAspGlyThrValTyrProGluLysGlyArgLeuLeuPheAlaAspProVal 273
Db 3022 CTCATCGAAGAACCGTCAACACTTCCCTCGAAGAGTACGCTGCAATTCCTCGATGAC 3081
QY 274 ValAsnGluSerThrGlyGlnIleThrLeuArgAlaAlaValProAsnAspGlnAsnIle 293
Db 3082 GTTGATGAAGAACCGCGCTTCATACCGTACGTCGTCTTCCCTAACCCGCAACATACG 3141
QY 294 LeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 3142 CTTTCCCGGCTATGTTTGTGCGTGAACGATGATGAAGCGCTCCAACTGACGCCATT 3201
QY 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetIleVal 331
Db 3202 CTCATCCCGCAACAGCGCTTACCGCACACCGCGGTGATGCAACCGTGCATGATTGTT 3261
QY 332 AsnAlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsn 351
Db 3262 AAGCATTAAGTAAGTCAGATGGAAGTCCCGCTGTCGTGCCAGCAGCGATGCGCATAA 3321
QY 352 TrpIleValThrSerLysLeuLysAspGlyAspLysValValAlaGlnGlyIleSerIle 371
Db 3322 TGGTATTAAGCAAGAGTCTGAATCTGGGATGCAAGTCACTTGTCAAGGCGCTG----- 3375
QY 372 AlaGlyIleThrGlyLysLysValThrProLysGlyThrAlaSerSerGluAsnGln 391
Db 3376 -----CAAAAGCGCGTCCGGAGAG----- 3396
QY 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db 3397 -----CAGTTAAAGCACTACGAGTAAACCCCGCAGATACTGATCGAAG 3441

```

```

RESULT 36
AP002564 307962 bp DNA linear BCT 07-MAR-2001
LOCUS AP002564
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 15/20.

```

[illegible]

```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative permease of ferrichrome ABC transporter"
/db_xref="GI:13363387"
/translaction="MSRIMVNGLRPLRVKRSSTLVLRKNTVLGLGFLFAVGLIFGL
MGSFFVPSVGVGALPAPRVSTDAIVYODIRLPVIVALLCGMIMGAGAMQSI
ARGADLPGLIGVKGCSVALMLIPQPMIGMFRVAVAGLGLVALVIFARSI
SRPFVLIGIVGSFPAAGIGVMTTADVDTVDVOTALLMGLSGSHAAWMLVGSACM
LPAALLLFTARADIALIGHVATGIVNUSRLALRVAPILITLVCSGVNIGR
VGLAIPHSRFLRGQOTTLILGSAVGLLIVADISGRILAPLQPLPAGIITISLIG
GPFFLLMQRRNSF"
/complement(3931..4914)
/gene="ECs3915"
/complement(3931..4914)
/gene="ECs3915"
/notice="probable ferrichrome ABC transporter
(permease), similar to ferrichrome ABC transporters
(permease) e.g. [Synecocystis sp.] g1/7442493|pir||574438
percent identity 43 in 315 aa. [Bacillus subtilis]
g1/1706795|sp|P4936|FHB_BACSU percent identity 39"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative permease of ferrichrome ABC transporter"
/protein_id="BAB37338.1"
/db_xref="GI:13363388"
/translaction="MMRTGLVTVIFLALLVGVVPGIARFIAPQTVLQAFLEHP
QNFHNIVRLRLPRLAALLTGSLVAGALLQAVRNPLGEPIILGLNAGLAV
AASALGIAPFVGRPLASTGALLFLILSLSSGRSLPMPKVLGLVALSPVSI
TALLIDBQTLAMRTMLAGDLAGDMATIGTSAWSLGGVLAITYLAPSLMALG
DMAQGVSVVSTRFTTLAIALLCANASINAGIPFVGLPQIVRIVSDLRVL
LPISACTGALLLADIIARITFTPHLAVMTALGAVFVIMATRMK"
/complement(4911..5720)
/gene="ECs3916"
/complement(4911..5720)
/gene="ECs3916"
/notice="probable ABC-type iron-siderophore transport system
ATP-binding protein, similar to ABC-type iron-siderophore
transport system ATP-binding proteins e.g. [Synecocystis
sp.] g1/7442509|pir||574440 percent identity 52 in 248 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ABC-type iron-siderophore transport
system ATP-binding protein"
/protein_id="BAB37339.1"
/db_xref="GI:13363389"
/translaction="MAHNAKQGLILDNVAGYHKKIIVDGVSPVPTKMTVLVGNAN
Alignment Scores:
Pred. No.: 2,98e-33 Length: 307962
Score: 749.50 Matches: 168
Percent Similarity: 60.65% Conservative: 74
Best Local Similarity: 42.11% Mismatches: 122
Query Match: 37.12% Indels: 35
DB: 1 Gaps: 6
US-09-889-756a-2 (1-412) x AP002564 (1-307962)
QY 17 AlalaValAlaValLeuSerSerCysGlyLyGlyValSpAlaAlaGlnGlyGln 36
Db 237572 TCCGCGCGCTTAAATCGCGGTGTAAAGATAGGAGAAAGAGCTCAGTCGGTGA 237631
QY 37 ProAlaGlyArgGluAlaProAlaProValValGlyValValThrValHis---ProGln 55
Db 237632 CCGCAG-----GTTACCGTTCATATTGTTAA 237658
QY 56 ThrValAlaLeuThrVal-----GluLeuProGlyValArgLeuGluSerLeuArgThrAla 73

```

```

Db 237659 ACGGCCCGCTTAGAAGTTAAAGTAACTGAATTACAGCGCCGACCAATGCTTATGTAAGCC 237718
QY 74 AspValArgAlaGlnValGlyGlyLeuLeuGlnLysArgLeuPheGlnGlySerTyr 93
Db 237719 GAAGTTCGCCACAGAGTGTACGGGATGCTGTAATGCAATTCTACTGAAGGACGAT 237778
QY 94 ValArgAlaGlyGlnProLeuTyrGlnLysAspSerSerThrTyrGluAlaAsnLeuGlu 113
Db 237779 GTCAACACAGCAGCAGTCCCTGTACAGATGATCCCGACCTATACAGCAAAATTATAC 237838
QY 114 SerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLysAlaAspLeu 133
Db 237839 AGCGCAAGACCGAAGCTGGGAAAGTGAAGCCGCCGCCCATGGCATGGATGACGTA 237898
QY 134 AlaArgTyrLysProLeuValAlaAlaGluAlaValSerArgGlnGluTyrAspAla 153
Db 237899 AAACGTTACGTTCCGCTCGGTGAGTACAAATATCATCGACAGAGTACGACAGCC 237958
QY 154 ValThrAlaLysArgSerAlaGlnAlaGlyValLysAlaAlaGlnAlaAlaLysSer 173
Db 237959 ATTGCTGATGCTCGCCAGCGCGATGCGCGCTGATTGCCCAAGACAGCATGGAAGC 238018
QY 174 AlaGlyLysLeuLeuAsnArgSerArgLeuThrAlaProLysSerGlyPheLeuGln 193
Db 238019 GCTCGCATCAATCTTCTTATACCAAGTACCTCGCCCAATTAGCGGACGATCGGCA 238078
QY 194 SerLeuValSerGluGlyThrLeuLeuAsnAlaGlyValSerThrValLeuAlaThrLe 213
Db 238079 TCACACTGACCAAGAGCGCTTGTCTACTAATGGCAACGATGACCTGACACTGTC 238138
QY 214 ArgGlnThrAsnProMetTyrValAsnValThrGlnSerAlaSerGluValMetLysLeu 233
Db 238139 CACAGCTGATCTCTATCTATGTTGATGATGACCAATCAGCAACGATTTATAGAGCTG 238198
QY 234 ArgArgGlnLeuAlaGlnGlyLysLeuLeuAlaAlaAspGlyValLeuAlaGlyLe 253
Db 238199 AACACAGCTCGTAAAGCAAGAAATTTGCATAGAGAAACCCCAACCAAGTAAAGT 238258
QY 254 LysPheAspArgGlyThrValTyrProGluLysGlyValArgLeuLeuPheAlaAspProVal 273
Db 238259 GTCATGGAAGAGCTCAAACTTATCCCTGAAGGTACGTTGCAATTCTCGATGTGACC 238318
QY 274 ValAsnGlnSerThrGlyGlnLeuThrLeuValAlaAlaValProAsnAspGlnAsnLe 293
Db 238319 GTTGATGAAGACACCGGCTCATACCTACGAGTGTCTTCCCTAACCGGCAACATACG 238378
QY 294 LeuMetProGlyLeuTyrValArgValLeuMetAspGlnValAlaValAspAsnAlaPhe 313
Db 238379 CTTTCCCGGGTATGTTTGGCGTGACGAGTATGAAAGCGCTCCAACTGACGCCATT 238438
QY 314 ValValProGlnGlnAlaValThrArgGlyAlaLysAsp-----ThrValMetLeuVal 331
Db 238439 CTATCCCGCAACAGCGCTTACCGCACACCGCGGTGATGCAACCGTCTGATTTGTT 238498
QY 332 AsnAlaGlnGlyLysMetGluProArgGluValThrValAlaGlnGlnGlnGlyThrAsn 351
Db 238499 AACGATMAAAGTACAGTGAAGTCCCGCTGTGTCGACGTACAGGAGATGGCGATAA 238558
QY 352 TrpIleValThrSerGlyLeuLysAspGlyAspLysValValValGlnGlyLysSerIle 371
Db 238559 TGGTGTATTAGCAAGATCTGAAATCTGGCGATCAAGTCAATGTCAGCGCGCTG----- 238612
QY 372 AlaGlyIleThrGlyLysLysValThrProLysGluThrAlaSerSerGluAsnGln 391
Db 238613 -----CAAAAGCGCGTCCGGAGAG----- 238633
QY 392 AlaAlaAlaProGlnSerGlyValGlnThrAlaSerGluAlaLysThrAlaSerGlu 410
Db 238634 -----CAGTTAAAGCACTACCGAATCCCGCAATATCTGATGGAAG 238678
RESULT 37
SMA252200 5914 bp DNA linear BCT 28-OCT-2000

```

DEFINITION Stenotrophomonas maltophilia smeD, smeB and smeF genes.
ACCESSION AJ252200
VERSION AJ252200.1 GI:11071582
KEYWORDS outer membrane protein; putative membrane fusion protein; putative
RND protein; smeD gene; smeE gene; smeF gene.
SOURCE Stenotrophomonas maltophilia
ORGANISM Stenotrophomonas maltophilia
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Stenotrophomonas.
REFERENCE
1 Alonso, A. and Martinez, J.L.
AUTHORS Cloning and characterization of SmeDEF, a novel multidrug efflux
TITLE pump from Stenotrophomonas maltophilia
JOURNAL Antimicrob. Agents Chemother. 44 (11), 3079-3086 (2000)
MEDLINE 20493115
PUBMED 11036026
REFERENCE 2 (bases 1 to 5914)
AUTHORS Martinez, J.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Martinez J., Departamento de Biotecnologia
Microbiana, Centro Nacional de Biotecnologia CSIC, Campus
Universitat Autonoma, Cerdoblanco, Madrid, 28049, SPAIN
Location/Qualifiers
FEATURES
source
1..5914
/organism="Stenotrophomonas maltophilia"
/mol_type="genomic DNA"
/isolate="D457R"
/db_xref="taxon:40324"
/note="D457R is a multiresistant mutant obtained by
single-step selection on tetracycline of D457"
82..1266
/gene="smeD"
82..1266
/gene="smeB"
/function="component of a multidrug efflux system"
/codon_start=1
/transl_table=1
/product="putative membrane fusion protein"
/protein_id="CAC14594.1"
/db_xref="GI:11071583"
/translation="MLSRLRPFALSLAIAAIVAAAGGQFQAPQEGDVTVTLKSE
TVGLRELPGRTNAFLVAEVRPQVNGIVAKRLFTGGQVAKGEPLYOLDASVYKQAN
NAAQLAAEATANAARLSAKRITELAKVDVSSQODLENAVAAKQKADGAKASL
DAAVTGVARITAPISGRIGKSVTQALVACQANALATVQOLDPIYDLTQSSAE
LLOLRLLAAGRLQDNOTPIVSTIMEDSGTFEHTLEFSVSVDPPTGSGKAVKD
NPDELKPMKTVRAVIGSGVRSADVLPQMGADPKCDITTAAMYVGDKNKRVKPVY
SRVGDKWLVDGKADKVIVEGLQIKIPGMVPAKTEKDPAPKAPPAAPAGD
AK"
1279..4401
/gene="smeB"
/gene="smeF"
1279..4401
/gene="smeB"
/gene="smeF"
/function="component of a multidrug efflux system"
/codon_start=1
/transl_table=1
/product="putative RND protein"
/protein_id="CAC14595.1"
/db_xref="GI:11071584"
/translation="MARFFIDRPIFAMVIAIILMAGGLAFKLPSVNPVAPAVE
ISATYPGASAKVEDSVTOILEONMKIGDLGIVSSSSSGQATITLTFSSGNPDI
AOVONKLIQAMPLLPOEVRQGINVAKSSGFLNAIAFSENGSDMANIDAYVS
NVVRLSPVGNIVQVGGKAMRIMLDPKLTYGLGVEVETLAKAQAVALAQ
LGAAPSVKGOOLNTINAOISRLTPEOPRNIIVRGADGRLRIGDVARVETLAESED
FVTRNGQPSAGLAVTLATGNALDPAAGVNALEDKGFPPAGLKAIEIPDTTPPR
VSIQVOTLLEALIVFVVMYLFLONRATLITLAVPVLLGTFGLAMLGFSVM
LTFMVAIALIGLVDALIVVENVERIMSEGLPELERTKSMQITGALIGLIVS
AVFPMAFMSGSTGVYIROFSATIVSAAALSVLVAIVLTPALCATMKPLKGSHVA
HRGLAGFPNGFNFGPRTSESQYGVGGLIHRPMRFGIVAAFLMGLVFLVPLPS
FLPNDSGVLMALVOAPVGAQOERTLESIALENHPLONEDADVASVSGSPFAG
GQNGAMAVKIKMSERDADNGVPIKRAMAALGQIKDAITPAFPAIPDELSTAG
YTFPLKNSGGHEALVAARQQLGLAASKLANVRNGEEDTPQFRLIDAKATS
LGSLIDQINGTLAAWSSSYIDFVDRKVRKRVQADQARWMPEDFQLSVNDKG
EMVPSAFATGKMDYGSPLERVNGVSMETQGBAPGVASGDMAETEQALQKLPAG

gene
4494..5894
/gene="smeF"
4494..5894
/gene="smeF"
/function="component of a multidrug efflux system"
/codon_start=1
/transl_table=1
/product="outer membrane protein"
/protein_id="CAC14596.1"
/db_xref="GI:11071585"
/translation="MEVIMPKSASLFSLATLALACSTLAPKNTAVAPAIQWPA
EAGGVADVAVAGMRDFTDARLQVETGSLQNNRDLRVAVLNVBARQVAVORAD
RNVGAVATGMDRGRDAGTBOFSAGVGAEEFLDFGRVRLSEALQVAVAN
RNAQLSLVAETATMWLTGADQRLKIDATLTKTIEDSLRLBAHRRGSSALBLT
QRTLVETATDARLQGLADRNALALAGGOLDPLPSIEQLLALPPAGL
PSDVLORPIMAAEHQLAANANIGAARAPPSISLTSIGSGSSLSNLPDSGR
VMSFLPKITLPIFGGKRLANLIALADRILAAOYEKSIQVGFRETDALNVSID
EQVSSQQRVLEAEQANRLSQARVYDAGLDSFTLLDARRAVNAQQTOLQAQAQAN
RITLYKVLGGMHERG"
BASE COUNT 1001 a 1979 c 1976 g 958 t
ORIGIN
Alignment Scores:
Pred. No.: 2,4e-35 Length: 5914
Score: 747.00 Matches: 170
Percent Similarity: 60.34% Conservative: 75
Best Local Similarity: 41.87% Mismatches: 145
Query Match: 37.00% Indels: 16
DB: 1 Gaps: 4
US-09-889-756a-2 (1-412) x SMA252200 (1-5914)
QY 9 MecArgAlaAlaAlaLeuAlaAlaValAlaLeuValLeuSerSerYsgLYVsgLY 28
Db 97 ATCCGACCCCTTGACAGTGCCTGCGCAATGCCGCGACCGCTGCTGC----- 147
QY 29 GlyAspAlaAlaGlnGlyGlnProAlaGlyArgGlnAlaProAlaProValGly 48
Db 148 -----GCCGCCCAACCCAGGCCGCCGAGAGGCGCGGTGACGTACC 192
QY 49 ValValThrValHisProGlnThrValAlaLeuThrValGluLeuProGlyArgLeuGlu 68
Db 193 GTGGTACGCTGAAGTCCGACCGTGGCTGACCCGCCAAGCTCGCGCGCGTACCAAT 252
QY 69 SerLeuArgThrAlaAspValArgAlaGlnValGlyIleIleGlnValArgLeuPhe 88
Db 253 GCCTTCCTGGTCCGCAAGTGGCGCCGACGTCAATGCAATCGGCGCAAGCGCTGTT 312
QY 89 GlnGlnGlySerThrValArgAlaGlyGlnProLeuThrGlnIleAspSerSerThrTyr 108
Db 313 ACCGAGGCGCGATGCTCAAGCGCGGAGACCGCTGACAGCTGACGATGCCAGCTAC 372
QY 109 GlnAlaAsnLeuGlnSerAlaArgAlaGlnLeuAlaThrAlaGlnAlaThrLeuAlaLys 128
Db 373 CGGCGCCAGGCCAACAGCGCCGCGACGTGCGCGCCGCAAGCCAGCCCAAGGCC 432
QY 129 AlaAspAlaAspLeuAlaArgTyrLysProLeuValAlaAlaGlnAlaValSerArgGln 148
Db 433 GCGCGCTGAGTGCACAGCGCATCACCGATGCCCAAGTGCATGCGGTGCAGCCACAG 492
QY 149 GlnTyrAspAlaAlaValThrAlaLysArgSerAlaGlnValLysAlaAlaGln 168
Db 493 GACCTGAGAACCGCGTCCGCGCAACAGCGCCAGCCAGCCAGCTGCGTCCGCGCAAG 552
QY 169 AlaAlaIleLysSerAlaGlyIleAsnLeuAsnArgSerArgIleThrAlaProIleSer 188
Db 553 GCCTCGGTGATGCGGCGCAACGTCACCTGAGCTACCGCGCATACCGCGCGCATACG 612
QY 189 GlyPheIleGlyGlnSerLysValSerGlnGlyThrLeuLeuAsnAlaGlyAspThrThr 208

Qy 174 Alaglyleasnlleuasnaraserglilethralaprolleserglypheilleglyln 193
 Db 215742 GCTTCGATCATCTCGCTTAAACCAAGTGAATCGCGCAATTACCGGACGATTCGGCAAA 215801
 Qy 194 Serlystalsergllylthrtleuasnlalaglyasprthrtthralleuallthille 213
 Db 215802 TCGACCGTGAACCGAGGCTCTTGTCATTAATGGCAACAGCTGAACCTGGCACTGTC 215861
 Qy 214 Argglnthrasnprometlyrvalasnlvalthrglnseralsasergluvalmetyleu 233
 Db 215862 CAGAGCTGCATCTTACTACTGATGTGACCCCAATCCAGCAACGATTTATAGGCTG 215921
 Qy 234 Argarglnllealaglylylveleuallalaspolyvalillealvalglyle 253
 Db 215922 AAGCAATCCGTAGGCAAGAAATTTGATAGGAAACCGCACAGCAAGTGAAGTTG 215981
 Qy 254 Lysphaspaaspglylthrtvaltyrprogllylthrtleuaphealasprowal 273
 Db 215982 GTCATGAGAACGGTCAACCTATCCCTGAAAGTACGCTGCAATTCCTCGATGTACC 216041
 Qy 274 Valasgluserthrglylgnllethrtleuasnlalavalproasaspglnasnlle 293
 Db 216042 GTTGATGAAGACCGGCTCATACCTTACCTGCTCTTCTTCCATACCGCAACATACG 216101
 Qy 294 Leuwerproglyleuylrvalargvalleuweraspnlalalalaspasnlalphe 313
 Db 216102 CTTTTCGGGTATGTTGTCGGCGACGATGATGAAAGCGCTCAACCTGACCATTCATT 216161
 Qy 314 Valvalproglnglnalavalthrtarglyalalyasp-----thrtvalmetilleval 331
 Db 216162 CTTATCCCGCAACAGCGGTTAGCCGACACACAGCTGTGATGCAACCGCTCATATGTT 216221
 Qy 332 Asnlaglnllylmetgluproarglvalthrtvalalaglnlgnllythrtasn 351
 Db 216222 AACCAATAAAGCTGCTTGAAGCCGCCCTGCTCTTCCAGCTCAGCGGATTCGGCAATAA 216281
 Qy 352 Trpillevalthrtserglyleuylaspolyaspilyalvalvalglylyleserile 371
 Db 216282 TGGTTGATTAAGCGAAGCTGAATCTGGCGATCAAGTCATTTGACGGGCTG----- 216335
 Qy 372 Alaglylethrtgylalylylvalthrtprolysglthrtalasersegluasngln 391
 Db 216336 -----CAAAAAGCGCTCCGGAGAG----- 216356
 Qy 392 Alalalalaproglinserglylvalglnthrtalaserglualystrthralaserglu 410
 Db 216357 -----CAGTTAAAGCCATACCGATACCCCGCATCTGCGATCGAAG 216401
 RESULT 39
 AE011925/c 12478 bp DNA linear BCT 29-MAY-2002
 LOCUS Xanthomonas axonopodis pv. citri str. 306, section 303 of 469 of
 the complete genome.
 ACCESSION AE011925 AE008923
 VERSION AE011925.1 GI:21109129
 KEYWORDS
 SOURCE Xanthomonas axonopodis pv. citri str. 306
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 Xanthomonadaceae; Xanthomonas.
 1 (bases 1 to 12478)
 da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
 Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
 Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
 Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,
 Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
 Curisno-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
 Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
 Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
 Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
 Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
 Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
 Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,

TITLE Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,
 Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
 Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
 Kitajima,J.P.
 JOURNAL Comparison of the genomes of two Xanthomonas pathogens with
 MEDLINE differing host specificities
 PUBMED Nature 417 (6887), 459-463 (2002)
 REFERENCE 22022145
 12024217
 2 (bases 1 to 12478)
 da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
 Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
 Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
 Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,
 Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
 Curisno-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
 Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
 Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
 Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
 Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
 Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
 Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
 Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,
 Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
 Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
 Kitajima,J.P.
 TITLE Direct Submission
 JOURNAL Submitted (28-NOV-2001) Departamento de Biogenetica, Universidade de
 Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
 Brazil
 FEATURES
 source
 location/Qualifiers
 1..12478
 /organism="Xanthomonas axonopodis pv. citri str. 306"
 /mol_type="genomic DNA"
 /strain="306"
 /db_xref="taxon:190486"
 /note="pathovar: citri"
 complement (72..1103)
 /gene="mcca"
 /note="synonym: XAC2835"
 complement (72..1103)
 /gene="mcca"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /transl_table=1
 /codon_start=1
 /product="oxidoreductase"
 /protein_id="AA037680.1"
 /db_xref="GI:21109130"
 /translation="METRLIGRSGFVKPVVLGLAGTGGKPLFSAWGDTDVAAARM
 IDLCLAGLNLPTATVYSGASEDLGLQALGRRQVILSTGTGRLGSGPNDAGS
 RRLRLAVENCRLRLTDYLDLQLHAPDMTMEETLSTLDGLVARGKRYIGASNY
 SGQMLKSGQVADPAGHNRVNAVQTYISLGRVYEWELMPLGLIDGIGAVASPLWG
 RLTKGLRQGPLPQHSRLHATAAGPAIDARLFDIVLADALAAETGRVSPQALVM
 LLQRPVSTVLIGARDEVQLRENIAGVWSLTGDORARLDQVAVPEPPYPYPMRQ
 FAERSPVPL"
 complement (1260..1403)
 /gene="XAC2836"
 complement (1260..1403)
 /gene="XAC2836"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /transl_table=1
 /codon_start=1
 /product="WFS transporter"
 /protein_id="AA037681.1"
 /db_xref="GI:21109131"
 /translation="MTATQAGLVATPWAALLSAVGIGALMSVQIGRRRAAPLNTC
 QMG"
 complement (1476..2477)
 /gene="arab"
 /note="synonym: XAC2837"
 complement (1476..2477)

gene
CDS

/gene="arau"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon start=1
/transl_table=11
/product="WFS transporter"
/protein_id="AA037682.1"
/db_xref="GI:21109132"
/translation="MSNLTARSRRPPALATLTGAFGIGTTEFVIMGLLOQVADL
GVSLASGILISGVALGVGAPVLTASARLPKAVLGMILFTVGNVACALAPY
TSLMAYVITSLAGTFPGVAVATSLVPRERASISLMFAGLTATLTCGPAWM
LGLOGRATFWAAIGVATSAVAWVPAAGATPVSRQSAVALQSGVLLAIA
ITVYGVAFVAFVYIQPLQVTFQSTVSPVLVFGVMTVGNLGGRLDRPT
AALGSLAALVVGALGCVLHSAKAAVTFGLGVAAFAVAPLQLRLEHARAG"
2577..3491
/gene="XAC2838"
2577..3491
/gene="XAC2838"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon start=1
/transl_table=11
/product="transcriptional regulator lysr family"
/protein_id="AA037683.1"
/db_xref="GI:21109133"
/translation="MDRLGDIPLFLRVLDGSIITAGARSIDLVAVASQRLKLEREL
GVRLHRTTRHLPTAGORLAERGVLEDDLAEDLESADVGVLKMTSASF
GROYVSLPFPQARHPRLRISADLDIVDQGLAIRGLEDGVLARIAD
NRRTLASPAYLOHGTPTPTDELQAAGVILMDRQDQVWTQTQDGLVRRMS
RFESNFBELVDAVAGGIAVHSYWHIADLRGRLVHWDPYPPASQISAVPAR
QLQPPRVAFVDFLOERLGEOPWELAT"
4196..5149
/gene="XAC2839"
4196..5149
/gene="XAC2839"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA037684.1"
/db_xref="GI:21109134"
/translation="MPSRLRIAPSTGCCMSLTSTLLRHAPGDMPTQIAGIQ
IMRSASPVTAMPYTPMLCIAQGRQAMGAQAYFHPQMYLAVADLPYGSVE
ASGOPYLFCFLDLDLTVSELAINHPELVDTORDASAGILNRSIPQDADLR
RLDOPHEIALAPLIVRELIXRLMADPANAAROMAIADSRINOISRAIYVREHA
ORESTEOIADLSMSRSTFNAHFKAVTTPFDYRSQLRHEARLWADALTTADAG
FRVYESASQPSRDYARILGISPARDAQRLANADATEAA"
complement (5264..5656)
/gene="XAC2840"
complement (5264..5656)
/gene="XAC2840"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA037685.1"
/db_xref="GI:21109135"
/translation="MSRAGAPSVVSAGYRTGKVRIDAPQADAPARVAGATVTEP
GARTMTHPLGOTLIVAGAGVQEWGKPAQOIRPDIVIPGVKHMIGANVANM
THIAVESODAPVTWLEOVSEAOYAAE"
complement (5831..6505)
/gene="XAC2841"
complement (5831..6505)
/gene="XAC2841"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA037686.1"
/db_xref="GI:21109136"

/translation="MDDAPPIARAPDHKGAIIAARVLFQGHGFDRTSMDTIAERA
MSKATVYAHFASKVELFRTTLEALQASPNRTALLAQPLEORLAAVADLVRS
ASMRBDAAVGLVRPPLPSOMREMMWTLCEFEYDPMRTLLREYORGALVIDNPD
ASVHFFGLMTGRATAARDADARGASVQDAVVSAGVALFLAAYRPDAVPGDSRSR
ELPSEF"
complement (6498..8009)
/gene="XAC2842"
/note="synonym: XAC2842"
complement (6498..8009)
/gene="XAC2842"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AA037687.1"
/db_xref="GI:21109137"
/translation="WTTIRTYCALVAVSLSECTIMPRYTRPDAPVPAQFADTVSV
APPAAATADTGNADTAIADIQWQVFPDPTLQOVIALALDNNDLRVALNTEVA
RAQYRVDRAALPAIDGTGNANARPELAIPOQVPTYSANIGISVYELDFOR
VRSLEQALQOQPLSTAEARSTHISVAEVAAYTLTAADQQLQAQSTQSDSY
RLQORSEBELGVASQTLTROKQTVETARVDEYRTQVODRNALVILVTOVPELL
PHALPNASVEGNVLASVAPAGLSQLRPPLLEARNLRANANIGARAFPEBSI
SLTASGSSSSSLRLFDGTRKMSFTVTLTLPFRAGRRANLDMKARNDEVARY
EKSIQAFREVSALAQRPDLGRLQAQALVATADTSRLSQARERGIDSYLQALD
AQSLVSAQOQTLINTQLSRFTNLVTFYKMGGMQTAAPAVASPAABPPRG"
complement (8011..11154)
/gene="XAC2843"
/note="synonym: XAC2843"
complement (8011..11154)
/gene="XAC2843"

Alignment Scores:
Pred. No.: 1,36e-34 Length: 12478
Score: 741.50 Matches: 174
Percent Similarity: 58.15% Conservative: 65
Best Local Similarity: 42.34% Mismatches: 151
Query Match: 36.73% Indels: 21
DB: 1 Gaps: 7

US-09-889-756a-2 (1-412) x AB011925 (1-12478)

Qy	8	Alawetarglaalaaalaleuaalaaalavala-----leuValleuSerSerCys	25
Db	12369	TCGTTTGCGCTGCGCTGCTGCTGCTGCTGCTGCGCCACCGCTTGTACGGCCTGC	12310
Qy	26	GIYLYS-----GIYLYSAPALAAAGLNGLYGILINPROALAGLYAGLUALAPro	43
Db	12309	GCGAGCGCGCGGAGCGCGCGCGCCACAGAAAGC-----	12274
Qy	44	Alaprovalalglyvalalthrvalhileproglinthrvalalaleuthrvalgluleu	63
Db	12273	ACCCGGAATGGCGCGTGTACCGTGAACAGACACCGGTGACGCTACACGACTTG	12214
Qy	64	ProgllyargleugluserleuarthrAlaaspvalarglaaglnvalglyglyleile	83
Db	12213	CCGGGGGCGACCGCCGCTGATCTCCGAATTCGCCGCAAGTGGAGGAGCATGTC	12154
Qy	84	GlnlyargleuhebngluclyserTyrValarglaaglylnproleuTyrGlnlle	103
Db	12153	CAGTCGCGCAGTTCACCGAAGCGCGGAGCGTCAAGCTGGCCAGACCTGTACCAATC	12094
Qy	104	AspserSerThrTyrGlnualaanleuGluserAlaarglaaglnleualatrraagln	123
Db	12093	GATCCGCGCAGTACCGCGCAGCTATGCGAGCGCGCAGCAATCCCTTGCAGAGCGGAG	12034
Qy	124	AlaThrleualalysalaaaplaaspaleualargTyrlyspProleuValalaaaglu	143
Db	12033	GCGACCTCGCGACCGCGGACGTCAGAGCGGAGCGGCTCAAGAGCGCGGAGCATCAA	11974
Qy	144	AlaValserargGlnGluTyrAspAlaAlaValThrAlaYsarSerAlaagluagly	163
Db	11973	GCGATCAGCGCAGAGGAGCGGACGACCGATGCCGCTAGCGCCAGCGCAGCGGAC	11914

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:09:43 ; Search time 25 Seconds

(without alignments)
2262.764 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412
Sequence: 1 MAFAFKAMRAAALAAVAL.....AAPSGVOTASEAKTASEAE 412Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 513375 seqs, 137303645 residues

Word size: 0

Total number of hits satisfying chosen parameters: 513375

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: Published Applications_Aa.*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_US_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.4	124	US-10-012-542-398	Sequence 398, App
2	10	2.4	222	US-09-890-688-154	Sequence 154, App
3	8	1.9	102	US-10-156-761-10004	Sequence 10034, A
4	8	1.9	181	US-10-080-170-577	Sequence 577, App
5	8	1.9	194	US-10-012-542-397	Sequence 397, App
6	8	1.9	478	US-09-813-555-1	Sequence 1, Appl1
7	8	1.9	478	US-09-813-555-2	Sequence 2, Appl1
8	8	1.9	478	US-09-813-555-3	Sequence 3, Appl1
9	8	1.9	478	US-09-943-123-6	Sequence 6, Appl1
10	8	1.9	478	US-09-943-123-7	Sequence 7, Appl1
11	8	1.9	478	US-09-943-123-12	Sequence 12, Appl1
12	8	1.9	478	US-09-943-123-13	Sequence 13, Appl1
13	8	1.9	480	US-09-943-123-16	Sequence 16, Appl1
14	8	1.9	480	US-09-943-123-17	Sequence 17, Appl1
15	8	1.9	500	US-10-106-698-4765	Sequence 4765, App

16	8	1.9	891	US-09-815-242-10147	Sequence 10147, A
17	8	1.9	892	US-09-815-242-13843	Sequence 13843, A
18	7	1.7	17	US-09-864-761-36687	Sequence 36687, A
19	7	1.7	24	US-09-983-802-347	Sequence 347, App
20	7	1.7	37	US-09-864-761-44448	Sequence 44448, App
21	7	1.7	65	US-09-764-877-1958	Sequence 1958, App
22	7	1.7	117	US-10-156-761-13107	Sequence 13107, A
23	7	1.7	126	US-10-007-267-2	Sequence 2, Appl1
24	7	1.7	132	US-09-956-425-4	Sequence 4, Appl1
25	7	1.7	132	US-09-867-550-150	Sequence 150, App
26	7	1.7	153	US-09-925-100-1688	Sequence 1688, App
27	7	1.7	175	US-10-156-761-10928	Sequence 10928, A
28	7	1.7	184	US-10-156-761-10784	Sequence 10784, A
29	7	1.7	203	US-09-764-853-558	Sequence 558, App
30	7	1.7	236	US-09-738-626-6552	Sequence 6552, App
31	7	1.7	248	US-10-156-761-13873	Sequence 13873, A
32	7	1.7	253	US-10-317-806-4	Sequence 4, Appl1
33	7	1.7	262	US-10-156-761-14170	Sequence 14170, A
34	7	1.7	271	US-10-342-224-72	Sequence 72, Appl1
35	7	1.7	275	US-09-815-242-10842	Sequence 10842, A
36	7	1.7	275	US-10-080-170-328	Sequence 328, App
37	7	1.7	288	US-10-156-761-14182	Sequence 14182, A
38	7	1.7	300	US-10-156-761-8083	Sequence 8083, App
39	7	1.7	301	US-10-156-761-13108	Sequence 13108, App
40	7	1.7	302	US-09-533-029-28	Sequence 28, Appl1
41	7	1.7	302	US-09-934-455-146	Sequence 146, App
42	7	1.7	302	US-10-295-403-86	Sequence 86, Appl1
43	7	1.7	302	US-10-278-536-56	Sequence 56, Appl1
44	7	1.7	304	US-10-156-761-14293	Sequence 14293, A
45	7	1.7	305	US-10-156-761-8219	Sequence 8219, App
46	7	1.7	320	US-09-738-626-4269	Sequence 4269, App
47	7	1.7	320	US-09-738-626-5385	Sequence 5385, App
48	7	1.7	339	US-09-802-371-2	Sequence 2, Appl1
49	7	1.7	351	US-10-156-761-17774	Sequence 7774, App
50	7	1.7	361	US-10-114-893-88	Sequence 88, Appl1
51	7	1.7	373	US-10-186-886-42	Sequence 42, Appl1
52	7	1.7	373	US-10-186-886-43	Sequence 43, Appl1
53	7	1.7	383	US-10-156-761-9603	Sequence 9603, App
54	7	1.7	396	US-09-712-363-231	Sequence 231, App
55	7	1.7	397	US-10-156-761-8560	Sequence 8560, App
56	7	1.7	434	US-10-102-806-506	Sequence 506, App
57	7	1.7	471	US-10-156-761-8883	Sequence 8883, App
58	7	1.7	471	US-10-156-761-14735	Sequence 14735, A
59	7	1.7	482	US-09-746-783-66	Sequence 66, Appl1
60	7	1.7	487	US-10-137-870-528	Sequence 528, App
61	7	1.7	487	US-10-140-018-528	Sequence 528, App
62	7	1.7	487	US-10-140-021-528	Sequence 528, App
63	7	1.7	487	US-10-140-021-528	Sequence 528, App
64	7	1.7	487	US-10-140-021-528	Sequence 528, App
65	7	1.7	487	US-10-140-021-528	Sequence 528, App
66	7	1.7	487	US-10-140-021-528	Sequence 528, App
67	7	1.7	487	US-10-140-021-528	Sequence 528, App
68	7	1.7	487	US-10-140-021-528	Sequence 528, App
69	7	1.7	487	US-10-141-698-528	Sequence 528, App
70	7	1.7	487	US-10-141-702-528	Sequence 528, App
71	7	1.7	487	US-10-141-704-528	Sequence 528, App
72	7	1.7	487	US-10-142-421-528	Sequence 528, App
73	7	1.7	487	US-10-142-432-528	Sequence 528, App
74	7	1.7	487	US-10-142-767-528	Sequence 528, App
75	7	1.7	487	US-10-143-033-528	Sequence 528, App
76	7	1.7	487	US-10-144-994-528	Sequence 528, App
77	7	1.7	487	US-10-145-528-528	Sequence 528, App
78	7	1.7	487	US-10-145-528-528	Sequence 528, App
79	7	1.7	487	US-10-145-528-528	Sequence 528, App
80	7	1.7	487	US-10-145-528-528	Sequence 528, App
81	7	1.7	487	US-10-145-528-528	Sequence 528, App
82	7	1.7	487	US-10-145-528-528	Sequence 528, App
83	7	1.7	487	US-10-145-528-528	Sequence 528, App
84	7	1.7	487	US-10-145-528-528	Sequence 528, App
85	7	1.7	487	US-10-145-528-528	Sequence 528, App
86	7	1.7	487	US-10-145-528-528	Sequence 528, App
87	7	1.7	487	US-10-146-724-528	Sequence 528, App
88	7	1.7	487	US-10-146-725-528	Sequence 528, App

89	7	1.7	487	12	US-10-146-795-528	Sequence 528, App	162	7	1.7	487	15	US-10-033-435-18	Sequence 18, App1
90	7	1.7	487	12	US-10-147-495-528	Sequence 528, App	163	7	1.7	487	15	US-10-175-746-528	Sequence 528, App
91	7	1.7	487	12	US-10-147-501-528	Sequence 528, App	164	7	1.7	487	15	US-10-176-918-528	Sequence 528, App
92	7	1.7	487	12	US-10-147-504-528	Sequence 528, App	165	7	1.7	487	15	US-10-176-921-528	Sequence 528, App
93	7	1.7	487	12	US-10-147-506-528	Sequence 528, App	166	7	1.7	487	15	US-10-227-884-204	Sequence 204, App
94	7	1.7	487	12	US-10-147-509-528	Sequence 528, App	167	7	1.7	487	15	US-10-032-990-18	Sequence 18, App1
95	7	1.7	487	12	US-10-147-510-528	Sequence 528, App	168	7	1.7	487	15	US-10-137-665-528	Sequence 528, App
96	7	1.7	487	12	US-10-147-511-528	Sequence 528, App	169	7	1.7	487	15	US-10-140-474-528	Sequence 528, App
97	7	1.7	487	12	US-10-147-512-528	Sequence 528, App	170	7	1.7	487	15	US-10-142-431-528	Sequence 528, App
98	7	1.7	487	12	US-10-147-513-528	Sequence 528, App	171	7	1.7	487	15	US-10-143-114-528	Sequence 528, App
99	7	1.7	487	12	US-10-152-397-528	Sequence 528, App	172	7	1.7	487	15	US-10-230-163-204	Sequence 204, App
100	7	1.7	487	12	US-10-153-386-528	Sequence 528, App	173	7	1.7	487	15	US-10-002-528	Sequence 528, App
101	7	1.7	487	12	US-10-158-783-528	Sequence 528, App	174	7	1.7	487	15	US-10-230-338-528	Sequence 204, App
102	7	1.7	487	12	US-10-158-786-528	Sequence 528, App	175	7	1.7	487	15	US-10-142-419-528	Sequence 528, App
103	7	1.7	487	12	US-10-140-019-528	Sequence 528, App	176	7	1.7	487	15	US-10-218-631-204	Sequence 204, App
104	7	1.7	487	12	US-10-140-022-528	Sequence 528, App	177	7	1.7	487	15	US-10-123-262-528	Sequence 528, App
105	7	1.7	487	12	US-10-140-861-528	Sequence 528, App	178	7	1.7	487	15	US-10-142-423-528	Sequence 528, App
106	7	1.7	487	12	US-10-140-862-528	Sequence 528, App	179	7	1.7	487	15	US-10-142-423-528	Sequence 528, App
107	7	1.7	487	12	US-10-141-697-528	Sequence 528, App	180	7	1.7	487	15	US-10-230-414-204	Sequence 204, App
108	7	1.7	487	12	US-10-141-700-528	Sequence 528, App	181	7	1.7	487	15	US-10-032-996-18	Sequence 18, App1
109	7	1.7	487	12	US-10-141-705-528	Sequence 528, App	182	7	1.7	487	15	US-10-121-050-528	Sequence 528, App
110	7	1.7	487	12	US-10-141-753-528	Sequence 528, App	183	7	1.7	487	15	US-10-141-755-528	Sequence 528, App
111	7	1.7	487	12	US-10-141-758-528	Sequence 528, App	184	7	1.7	487	15	US-10-143-032-528	Sequence 528, App
112	7	1.7	487	12	US-10-142-418-528	Sequence 528, App	185	7	1.7	487	15	US-10-123-108-528	Sequence 528, App
113	7	1.7	487	12	US-10-142-420-528	Sequence 528, App	186	7	1.7	487	15	US-10-123-236-528	Sequence 528, App
114	7	1.7	487	12	US-10-142-422-528	Sequence 528, App	187	7	1.7	487	15	US-10-123-261-528	Sequence 528, App
115	7	1.7	487	12	US-10-142-427-528	Sequence 528, App	188	7	1.7	487	15	US-10-140-921-528	Sequence 528, App
116	7	1.7	487	12	US-10-143-760-528	Sequence 528, App	189	7	1.7	487	15	US-10-140-928-528	Sequence 528, App
117	7	1.7	487	12	US-10-145-821-528	Sequence 528, App	190	7	1.7	487	15	US-10-216-159A-204	Sequence 204, App
118	7	1.7	487	12	US-10-152-531-528	Sequence 528, App	191	7	1.7	487	15	US-10-121-045-528	Sequence 528, App
119	7	1.7	487	12	US-10-216-163-204	Sequence 204, App	192	7	1.7	487	15	US-10-123-282-528	Sequence 528, App
120	7	1.7	487	12	US-10-127-840A-528	Sequence 528, App	193	7	1.7	487	15	US-10-123-903-528	Sequence 528, App
121	7	1.7	487	12	US-10-143-761-528	Sequence 528, App	194	7	1.7	487	15	US-10-124-819-528	Sequence 528, App
122	7	1.7	487	12	US-10-143-763-528	Sequence 528, App	195	7	1.7	487	15	US-10-140-925-528	Sequence 528, App
123	7	1.7	487	12	US-10-142-765-528	Sequence 528, App	196	7	1.7	487	15	US-10-140-925-528	Sequence 528, App
124	7	1.7	487	12	US-10-142-887-528	Sequence 528, App	197	7	1.7	487	15	US-10-160-498-528	Sequence 528, App
125	7	1.7	487	12	US-10-142-888-528	Sequence 528, App	198	7	1.7	487	15	US-10-218-449-528	Sequence 204, App
126	7	1.7	487	12	US-10-143-034-528	Sequence 528, App	199	7	1.7	487	15	US-10-227-873-204	Sequence 204, App
127	7	1.7	487	12	US-10-143-116-528	Sequence 528, App	200	7	1.7	487	15	US-10-227-883-528	Sequence 18, App1
128	7	1.7	487	12	US-10-144-957-528	Sequence 528, App	201	7	1.7	487	15	US-10-033-396-18	Sequence 18, App1
129	7	1.7	487	12	US-10-144-952-528	Sequence 528, App	202	7	1.7	487	15	US-10-124-824-528	Sequence 528, App
130	7	1.7	487	12	US-10-145-015-528	Sequence 528, App	203	7	1.7	487	15	US-10-127-825A-528	Sequence 528, App
131	7	1.7	487	12	US-10-145-015-528	Sequence 528, App	204	7	1.7	487	15	US-10-127-829A-528	Sequence 528, App
132	7	1.7	487	12	US-10-145-029-528	Sequence 528, App	205	7	1.7	487	15	US-10-127-835A-528	Sequence 528, App
133	7	1.7	487	12	US-10-145-091-528	Sequence 528, App	206	7	1.7	487	15	US-10-127-839A-528	Sequence 528, App
134	7	1.7	487	12	US-10-145-629-528	Sequence 528, App	207	7	1.7	487	15	US-10-127-901A-528	Sequence 528, App
135	7	1.7	487	12	US-10-145-630-528	Sequence 528, App	208	7	1.7	487	15	US-10-128-693A-528	Sequence 528, App
136	7	1.7	487	12	US-10-145-747-528	Sequence 528, App	209	7	1.7	487	15	US-10-131-813A-528	Sequence 528, App
137	7	1.7	487	12	US-10-145-754-528	Sequence 528, App	210	7	1.7	487	15	US-10-131-818A-528	Sequence 528, App
138	7	1.7	487	12	US-10-145-754-528	Sequence 528, App	211	7	1.7	487	15	US-10-131-823A-528	Sequence 528, App
139	7	1.7	487	12	US-10-145-755-528	Sequence 528, App	212	7	1.7	487	15	US-10-131-824A-528	Sequence 528, App
140	7	1.7	487	12	US-10-145-818-528	Sequence 528, App	213	7	1.7	487	15	US-10-131-827A-528	Sequence 528, App
141	7	1.7	487	12	US-10-145-820-528	Sequence 528, App	214	7	1.7	487	15	US-10-137-828A-528	Sequence 528, App
142	7	1.7	487	12	US-10-145-872-528	Sequence 528, App	215	7	1.7	487	15	US-10-147-500-528	Sequence 528, App
143	7	1.7	487	12	US-10-145-873-528	Sequence 528, App	216	7	1.7	487	15	US-10-147-502-528	Sequence 528, App
144	7	1.7	487	12	US-10-147-481-528	Sequence 528, App	217	7	1.7	487	15	US-10-147-502-528	Sequence 528, App
145	7	1.7	487	12	US-10-147-482-528	Sequence 528, App	218	7	1.7	487	15	US-10-147-515-528	Sequence 528, App
146	7	1.7	487	12	US-10-147-503-528	Sequence 528, App	219	7	1.7	487	15	US-10-147-517-528	Sequence 528, App
147	7	1.7	487	12	US-10-147-522-528	Sequence 528, App	220	7	1.7	487	15	US-10-147-526-528	Sequence 528, App
148	7	1.7	487	12	US-10-152-401-528	Sequence 528, App	221	7	1.7	487	15	US-10-147-527-528	Sequence 528, App
149	7	1.7	487	12	US-10-157-783-528	Sequence 528, App	222	7	1.7	487	15	US-10-121-041-528	Sequence 528, App
150	7	1.7	487	12	US-10-158-463-528	Sequence 528, App	223	7	1.7	487	15	US-10-121-043-528	Sequence 528, App
151	7	1.7	487	12	US-10-158-792-528	Sequence 528, App	224	7	1.7	487	15	US-10-121-047-528	Sequence 528, App
152	7	1.7	487	13	US-10-033-246-18	Sequence 18, App1	225	7	1.7	487	15	US-10-123-215-528	Sequence 528, App
153	7	1.7	487	13	US-10-033-301-18	Sequence 18, App1	226	7	1.7	487	15	US-10-123-902-528	Sequence 528, App
154	7	1.7	487	13	US-10-033-326-18	Sequence 18, App1	227	7	1.7	487	15	US-10-123-908-528	Sequence 528, App
155	7	1.7	487	14	US-10-033-245-18	Sequence 18, App1	228	7	1.7	487	15	US-10-123-909-528	Sequence 528, App
156	7	1.7	487	14	US-10-033-223-18	Sequence 18, App1	229	7	1.7	487	15	US-10-123-910-528	Sequence 528, App
157	7	1.7	487	14	US-10-033-167-18	Sequence 18, App1	230	7	1.7	487	15	US-10-124-813-528	Sequence 528, App
158	7	1.7	487	14	US-10-028-244-18	Sequence 18, App1	231	7	1.7	487	15	US-10-124-817-528	Sequence 528, App
159	7	1.7	487	15	US-10-028-072-528	Sequence 528, App	232	7	1.7	487	15	US-10-125-922-528	Sequence 528, App
160	7	1.7	487	15	US-10-121-049-528	Sequence 528, App	233	7	1.7	487	15	US-10-125-924-528	Sequence 528, App
161	7	1.7	487	15	US-10-123-904-528	Sequence 528, App	234	7	1.7	487	15	US-10-140-860-528	Sequence 528, App
	7	1.7	487	15	US-10-140-470-528	Sequence 528, App		7	1.7	487	15	US-10-142-417-528	Sequence 528, App

235	1.7	487	15	US-10-147-519-528	Sequence 528, App	308	7	1.7	487	15	US-10-131-817A-528	Sequence 528, App
236	1.7	487	15	US-10-157-782-528	Sequence 528, App	309	7	1.7	487	15	US-10-131-821A-528	Sequence 528, App
237	1.7	487	15	US-10-158-395-528	Sequence 528, App	310	7	1.7	487	15	US-10-131-822A-528	Sequence 528, App
238	1.7	487	15	US-10-219-076-204	Sequence 204, App	311	7	1.7	487	15	US-10-131-828A-528	Sequence 528, App
239	1.7	487	15	US-10-230-434-204	Sequence 528, App	312	7	1.7	487	15	US-10-131-835A-528	Sequence 528, App
240	1.7	487	15	US-10-125-926A-528	Sequence 528, App	313	7	1.7	487	15	US-10-137-869A-528	Sequence 528, App
241	1.7	487	15	US-10-125-930A-528	Sequence 528, App	314	7	1.7	487	15	US-10-137-869A-528	Sequence 528, App
242	1.7	487	15	US-10-127-831A-528	Sequence 528, App	315	7	1.7	487	15	US-10-147-523-528	Sequence 528, App
243	1.7	487	15	US-10-127-837A-528	Sequence 528, App	316	7	1.7	487	15	US-10-158-785-528	Sequence 528, App
244	1.7	487	15	US-10-127-838B-528	Sequence 528, App	317	7	1.7	487	15	US-10-121-051-528	Sequence 528, App
245	1.7	487	15	US-10-127-843A-528	Sequence 528, App	318	7	1.7	487	15	US-10-218-165-204	Sequence 204, App
246	1.7	487	15	US-10-127-843A-528	Sequence 528, App	319	7	1.7	487	15	US-10-218-956-204	Sequence 204, App
247	1.7	487	15	US-10-127-845A-528	Sequence 528, App	320	7	1.7	487	15	US-10-219-468-204	Sequence 204, App
248	1.7	487	15	US-10-127-846A-528	Sequence 528, App	321	7	1.7	487	15	US-10-219-478-204	Sequence 204, App
249	1.7	487	15	US-10-127-848A-528	Sequence 528, App	322	7	1.7	487	15	US-10-219-536-204	Sequence 204, App
250	1.7	487	15	US-10-127-849A-528	Sequence 528, App	323	7	1.7	487	15	US-10-233-205-204	Sequence 204, App
251	1.7	487	15	US-10-127-850A-528	Sequence 528, App	324	7	1.7	487	15	US-10-121-042-528	Sequence 528, App
252	1.7	487	15	US-10-127-851A-528	Sequence 528, App	325	7	1.7	487	15	US-10-219-072-204	Sequence 204, App
253	1.7	487	15	US-10-128-684A-528	Sequence 528, App	326	7	1.7	487	15	US-10-219-470-204	Sequence 204, App
254	1.7	487	15	US-10-128-686A-528	Sequence 528, App	327	7	1.7	487	15	US-10-219-474-204	Sequence 204, App
255	1.7	487	15	US-10-128-680A-528	Sequence 528, App	328	7	1.7	487	15	US-10-219-524-204	Sequence 204, App
256	1.7	487	15	US-10-128-691A-528	Sequence 528, App	329	7	1.7	487	15	US-10-219-528-204	Sequence 204, App
257	1.7	487	15	US-10-131-819A-528	Sequence 528, App	330	7	1.7	487	15	US-10-227-880-204	Sequence 204, App
258	1.7	487	15	US-10-131-829A-528	Sequence 528, App	331	7	1.7	487	15	US-10-227-881-204	Sequence 204, App
259	1.7	487	15	US-10-131-836A-528	Sequence 528, App	332	7	1.7	487	15	US-10-227-882-204	Sequence 204, App
260	1.7	487	15	US-10-146-729-528	Sequence 528, App	333	7	1.7	487	15	US-10-230-436-204	Sequence 204, App
261	1.7	487	15	US-10-146-791-528	Sequence 528, App	334	7	1.7	487	15	US-10-232-223-204	Sequence 204, App
262	1.7	487	15	US-10-147-484-528	Sequence 528, App	335	7	1.7	487	15	US-10-232-225-204	Sequence 204, App
263	1.7	487	15	US-10-147-508-528	Sequence 528, App	336	7	1.7	487	15	US-10-232-227-204	Sequence 204, App
264	1.7	487	15	US-10-147-512-528	Sequence 528, App	337	7	1.7	487	15	US-10-232-229-204	Sequence 204, App
265	1.7	487	15	US-10-175-735-528	Sequence 528, App	338	7	1.7	487	15	US-10-232-234-204	Sequence 204, App
266	1.7	487	15	US-10-121-040-528	Sequence 528, App	339	7	1.7	487	15	US-10-232-167-204	Sequence 204, App
267	1.7	487	15	US-10-131-056-528	Sequence 528, App	340	7	1.7	487	15	US-10-123-912-528	Sequence 528, App
268	1.7	487	15	US-10-121-061-528	Sequence 528, App	341	7	1.7	487	15	US-10-223-085-250	Sequence 250, App
269	1.7	487	15	US-10-123-235-528	Sequence 528, App	342	7	1.7	487	15	US-10-216-160-204	Sequence 204, App
270	1.7	487	15	US-10-124-815-528	Sequence 528, App	343	7	1.7	487	15	US-10-216-162-204	Sequence 204, App
271	1.7	487	15	US-10-137-868-528	Sequence 528, App	344	7	1.7	487	15	US-10-216-164-204	Sequence 204, App
272	1.7	487	15	US-10-147-492-528	Sequence 528, App	345	7	1.7	487	15	US-10-216-167-204	Sequence 204, App
273	1.7	487	15	US-10-158-782-528	Sequence 528, App	346	7	1.7	487	15	US-10-216-168-204	Sequence 204, App
274	1.7	487	15	US-10-123-905-528	Sequence 528, App	347	7	1.7	487	15	US-10-219-065-204	Sequence 204, App
275	1.7	487	15	US-10-124-815-528	Sequence 528, App	348	7	1.7	487	15	US-10-219-071-204	Sequence 204, App
276	1.7	487	15	US-10-125-921A-528	Sequence 528, App	349	7	1.7	487	15	US-10-219-074-204	Sequence 204, App
277	1.7	487	15	US-10-125-928A-528	Sequence 528, App	350	7	1.7	487	15	US-10-219-077-204	Sequence 204, App
278	1.7	487	15	US-10-127-821A-528	Sequence 528, App	351	7	1.7	487	15	US-10-219-465-204	Sequence 204, App
279	1.7	487	15	US-10-127-821A-528	Sequence 528, App	352	7	1.7	487	15	US-10-219-467-204	Sequence 204, App
280	1.7	487	15	US-10-127-822A-528	Sequence 528, App	353	7	1.7	487	15	US-10-219-469-204	Sequence 204, App
281	1.7	487	15	US-10-127-824A-528	Sequence 528, App	354	7	1.7	487	15	US-10-219-471-204	Sequence 204, App
282	1.7	487	15	US-10-127-826A-528	Sequence 528, App	355	7	1.7	487	15	US-10-219-473-204	Sequence 204, App
283	1.7	487	15	US-10-127-827A-528	Sequence 528, App	356	7	1.7	487	15	US-10-219-476-204	Sequence 204, App
284	1.7	487	15	US-10-127-828A-528	Sequence 528, App	357	7	1.7	487	15	US-10-219-482-204	Sequence 204, App
285	1.7	487	15	US-10-127-830A-528	Sequence 528, App	358	7	1.7	487	15	US-10-227-874-204	Sequence 204, App
286	1.7	487	15	US-10-127-832A-528	Sequence 528, App	359	7	1.7	487	15	US-10-227-876-204	Sequence 204, App
287	1.7	487	15	US-10-127-833A-528	Sequence 528, App	360	7	1.7	487	15	US-10-227-878-204	Sequence 204, App
288	1.7	487	15	US-10-127-834A-528	Sequence 528, App	361	7	1.7	487	15	US-10-229-974-204	Sequence 204, App
289	1.7	487	15	US-10-127-836A-528	Sequence 528, App	362	7	1.7	487	15	US-10-230-024-204	Sequence 204, App
290	1.7	487	15	US-10-127-841A-528	Sequence 528, App	363	7	1.7	487	15	US-10-230-113-204	Sequence 204, App
291	1.7	487	15	US-10-127-844A-528	Sequence 528, App	364	7	1.7	487	15	US-10-230-183-204	Sequence 204, App
292	1.7	487	15	US-10-128-687A-528	Sequence 528, App	365	7	1.7	487	15	US-10-230-234-204	Sequence 204, App
293	1.7	487	15	US-10-128-688A-528	Sequence 528, App	366	7	1.7	487	15	US-10-230-306-204	Sequence 204, App
294	1.7	487	15	US-10-128-689A-528	Sequence 528, App	367	7	1.7	487	15	US-10-230-426-204	Sequence 204, App
295	1.7	487	15	US-10-128-694A-528	Sequence 528, App	368	7	1.7	487	15	US-10-230-427-204	Sequence 204, App
296	1.7	487	15	US-10-131-825A-528	Sequence 528, App	369	7	1.7	487	15	US-10-230-433-204	Sequence 204, App
297	1.7	487	15	US-10-230-417-528	Sequence 204, App	370	7	1.7	487	15	US-10-230-435-204	Sequence 204, App
298	1.7	487	15	US-10-219-003-204	Sequence 204, App	371	7	1.7	487	15	US-10-230-438-204	Sequence 204, App
299	1.7	487	15	US-10-219-075-204	Sequence 204, App	372	7	1.7	487	15	US-10-232-222-204	Sequence 204, App
300	1.7	487	15	US-10-219-464-204	Sequence 204, App	373	7	1.7	487	15	US-10-192-007-528	Sequence 528, App
301	1.7	487	15	US-10-219-466-204	Sequence 204, App	374	7	1.7	487	15	US-10-194-359-528	Sequence 528, App
302	1.7	487	15	US-10-219-479-204	Sequence 204, App	375	7	1.7	487	15	US-10-223-084-250	Sequence 250, App
303	1.7	487	15	US-10-219-481-204	Sequence 204, App	376	7	1.7	487	15	US-10-223-088-250	Sequence 250, App
304	1.7	487	15	US-10-230-260-204	Sequence 204, App	377	7	1.7	487	15	US-10-223-090-250	Sequence 250, App
305	1.7	487	15	US-10-232-231-204	Sequence 204, App	378	7	1.7	487	15	US-10-219-070-204	Sequence 204, App
306	1.7	487	15	US-10-232-233-204	Sequence 204, App	379	7	1.7	487	15	US-10-219-527-204	Sequence 204, App
307	1.7	487	15	US-10-131-815A-528	Sequence 528, App	380	7	1.7	487	15	US-10-219-547-204	Sequence 204, App

381	7	1.7	487	15	US-10-227-877-204	Sequence 204, App	454	7	1.7	615	11	US-09-997-628-194	Sequence 194, App
382	7	1.7	487	15	US-10-223-087-250	Sequence 250, App	455	7	1.7	615	11	US-09-997-683-194	Sequence 194, App
383	7	1.7	487	15	US-10-127-8847A-528	Sequence 528, App	456	7	1.7	615	11	US-09-989-729A-194	Sequence 194, App
384	7	1.7	487	15	US-10-223-083-250	Sequence 250, App	457	7	1.7	615	11	US-09-997-449-194	Sequence 194, App
385	7	1.7	487	15	US-10-218-166-204	Sequence 204, App	458	7	1.7	615	11	US-09-997-440-194	Sequence 194, App
386	7	1.7	487	15	US-10-218-612-204	Sequence 204, App	459	7	1.7	615	11	US-09-990-440-194	Sequence 194, App
387	7	1.7	487	15	US-10-175-590-528	Sequence 528, App	460	7	1.7	615	11	US-09-993-469-194	Sequence 194, App
388	7	1.7	487	15	US-10-223-089-250	Sequence 250, App	461	7	1.7	615	11	US-09-997-442-194	Sequence 194, App
389	7	1.7	487	15	US-10-137-866-528	Sequence 528, App	462	7	1.7	615	11	US-09-993-748-194	Sequence 194, App
390	7	1.7	487	16	US-10-146-726-528	Sequence 528, App	463	7	1.7	615	11	US-09-990-439-194	Sequence 194, App
391	7	1.7	487	16	US-10-146-727-528	Sequence 528, App	464	7	1.7	615	11	US-09-990-427-194	Sequence 194, App
392	7	1.7	487	16	US-10-146-788-528	Sequence 528, App	465	7	1.7	615	11	US-09-989-328-194	Sequence 194, App
393	7	1.7	487	16	US-10-152-380-528	Sequence 528, App	466	7	1.7	615	11	US-09-993-883-194	Sequence 194, App
394	7	1.7	487	16	US-10-153-934-528	Sequence 528, App	467	7	1.7	615	11	US-09-941-992-194	Sequence 194, App
395	7	1.7	489	11	US-09-970-367-4	Sequence 4, Appl	468	7	1.7	615	11	US-09-992-521-194	Sequence 194, App
396	7	1.7	497	15	US-10-156-761-9418	Sequence 9418, App	469	7	1.7	615	11	US-09-997-333-194	Sequence 194, App
397	7	1.7	507	15	US-10-156-761-13988	Sequence 13988, App	470	7	1.7	615	11	US-09-997-384-194	Sequence 194, App
398	7	1.7	512	12	US-10-259-165-234	Sequence 234, App	471	7	1.7	615	11	US-09-998-041-194	Sequence 194, App
399	7	1.7	515	15	US-10-156-761-11280	Sequence 11280, App	472	7	1.7	615	11	US-09-997-585-194	Sequence 194, App
400	7	1.7	520	15	US-10-156-761-11280	Sequence 11280, App	473	7	1.7	615	11	US-09-997-614-194	Sequence 194, App
401	7	1.7	528	11	US-09-934-455-406	Sequence 406, App	474	7	1.7	615	12	US-09-989-662-194	Sequence 194, App
402	7	1.7	536	11	US-09-934-455-58	Sequence 58, App	475	7	1.7	615	12	US-09-989-725-194	Sequence 194, App
403	7	1.7	536	15	US-10-278-173-40	Sequence 40, Appl	476	7	1.7	615	12	US-09-997-529-194	Sequence 194, App
404	7	1.7	536	16	US-10-278-536-78	Sequence 78, Appl	477	7	1.7	631	15	US-10-156-880-16	Sequence 16, App
405	7	1.7	544	9	US-09-801-852A-4	Sequence 4, Appl	478	7	1.7	647	15	US-10-156-761-15036	Sequence 15036, App
406	7	1.7	550	15	US-10-156-761-8003	Sequence 8003, App	479	7	1.7	745	15	US-10-156-761-10715	Sequence 10715, App
407	7	1.7	570	9	US-09-874-923-104	Sequence 104, App	480	7	1.7	809	15	US-10-156-761-10945	Sequence 10945, App
408	7	1.7	570	9	US-09-991-496-104	Sequence 104, App	481	7	1.7	808	9	US-09-925-301-1161	Sequence 1161, App
409	7	1.7	572	9	US-09-801-852A-2	Sequence 2, Appl	482	7	1.7	946	9	US-09-712-163-213	Sequence 213, App
410	7	1.7	589	15	US-10-156-761-11277	Sequence 11277, App	483	7	1.7	959	11	US-09-927-827-61	Sequence 61, App
411	7	1.7	591	12	US-10-328-675A-66	Sequence 66, Appl	484	7	1.7	1078	15	US-10-058-124-21	Sequence 21, App
412	7	1.7	602	15	US-10-010-160-16	Sequence 16, Appl	485	7	1.7	1212	10	US-09-981-553-16	Sequence 16, App
413	7	1.7	609	15	US-10-156-761-10122	Sequence 10122, App	486	7	1.7	1212	12	US-10-241-220-80	Sequence 80, App
414	7	1.7	615	9	US-09-989-722-194	Sequence 194, App	487	7	1.7	1212	12	US-10-301-822-191	Sequence 191, App
415	7	1.7	615	9	US-09-989-723-194	Sequence 194, App	488	7	1.7	1212	15	US-10-176-847-14	Sequence 74, App
416	7	1.7	615	9	US-09-989-723-194	Sequence 194, App	489	7	1.7	1212	15	US-10-264-820-10	Sequence 10, App
417	7	1.7	615	9	US-09-989-727-194	Sequence 194, App	490	7	1.7	1230	15	US-10-156-761-14097	Sequence 14097, App
418	7	1.7	615	10	US-09-989-731-194	Sequence 194, App	491	7	1.7	1466	12	US-10-301-822-33	Sequence 33, App
419	7	1.7	615	10	US-09-989-732-194	Sequence 194, App	492	7	1.7	1466	15	US-10-177-293-68	Sequence 68, App
420	7	1.7	615	10	US-09-991-073-194	Sequence 194, App	493	7	1.7	1466	10	US-09-862-027-81	Sequence 81, App
421	7	1.7	615	10	US-09-990-442-194	Sequence 194, App	494	7	1.7	2365	15	US-10-756-761-14880	Sequence 14880, App
422	7	1.7	615	10	US-09-991-163-194	Sequence 194, App	495	7	1.7	2993	10	US-09-738-626-6239	Sequence 6239, App
423	7	1.7	615	10	US-09-993-604-194	Sequence 194, App	496	6	1.5	8	11	US-09-880-748-3142	Sequence 3142, App
424	7	1.7	615	10	US-09-990-456-194	Sequence 194, App	497	6	1.5	9	10	US-09-920-174-13	Sequence 17, App
425	7	1.7	615	10	US-09-989-721-194	Sequence 194, App	498	6	1.5	10	11	US-09-572-604B-1824	Sequence 1824, App
426	7	1.7	615	10	US-09-993-598-194	Sequence 194, App	499	6	1.5	18	15	US-10-056-884-19	Sequence 19, App
427	7	1.7	615	10	US-09-988-293A-194	Sequence 194, App	500	6	1.5	18	15	US-10-084-813-474	Sequence 474, App
428	7	1.7	615	10	US-09-988-735-194	Sequence 194, App	501	6	1.5	20	15	US-10-125-852-20	Sequence 20, App
429	7	1.7	615	10	US-09-990-444-194	Sequence 194, App	502	6	1.5	20	15	US-10-162-538-2	Sequence 2, Appl
430	7	1.7	615	10	US-09-991-181-194	Sequence 194, App	503	6	1.5	24	15	US-10-162-538-7	Sequence 7, Appl
431	7	1.7	615	10	US-09-989-730-194	Sequence 194, App	504	6	1.5	27	14	US-10-010-114-9	Sequence 9, Appl
432	7	1.7	615	10	US-09-990-436-194	Sequence 194, App	505	6	1.5	32	15	US-10-814-021-7	Sequence 7, Appl
433	7	1.7	615	10	US-09-993-687-194	Sequence 194, App	506	6	1.5	32	9	US-09-864-761-37273	Sequence 37273, App
434	7	1.7	615	11	US-09-988-734-194	Sequence 194, App	507	6	1.5	36	9	US-09-864-761-45637	Sequence 45637, App
435	7	1.7	615	11	US-09-997-653-194	Sequence 194, App	508	6	1.5	38	9	US-09-864-761-42372	Sequence 42372, App
436	7	1.7	615	11	US-09-993-667-194	Sequence 194, App	509	6	1.5	39	9	US-09-764-869-1050	Sequence 1050, App
437	7	1.7	615	11	US-09-997-428-194	Sequence 194, App	510	6	1.5	39	15	US-10-091-564-1050	Sequence 1050, App
438	7	1.7	615	11	US-09-997-666-194	Sequence 194, App	511	6	1.5	40	9	US-09-864-761-34563	Sequence 34563, App
439	7	1.7	615	11	US-09-990-438-194	Sequence 194, App	512	6	1.5	41	9	US-09-864-761-41183	Sequence 41183, App
440	7	1.7	615	11	US-09-990-562-194	Sequence 194, App	513	6	1.5	44	11	US-09-491-614-31	Sequence 31, Appl
441	7	1.7	615	11	US-09-990-711-194	Sequence 194, App	514	6	1.5	44	14	US-10-042-417-69	Sequence 69, Appl
442	7	1.7	615	11	US-09-988-726-194	Sequence 194, App	515	6	1.5	49	9	US-09-864-761-43234	Sequence 43234, App
443	7	1.7	615	11	US-09-998-156-194	Sequence 194, App	516	6	1.5	52	9	US-09-864-761-43647	Sequence 43647, App
444	7	1.7	615	11	US-09-990-437-194	Sequence 194, App	517	6	1.5	54	9	US-09-864-761-38359	Sequence 38359, App
445	7	1.7	615	11	US-09-991-157-194	Sequence 194, App	518	6	1.5	55	10	US-09-766-692-1533	Sequence 1533, App
446	7	1.7	615	11	US-09-997-514-194	Sequence 194, App	519	6	1.5	55	13	US-10-040-862-1533	Sequence 1533, App
447	7	1.7	615	11	US-09-997-573-194	Sequence 194, App	520	6	1.5	56	9	US-09-864-761-19143	Sequence 19143, App
448	7	1.7	615	11	US-09-991-172-194	Sequence 194, App	521	6	1.5	58	9	US-09-815-242-11764	Sequence 11764, App
449	7	1.7	615	11	US-09-990-726-194	Sequence 194, App	522	6	1.5	61	15	US-10-023-282-260	Sequence 260, App
450	7	1.7	615	11	US-09-997-559-194	Sequence 194, App	523	6	1.5	64	15	US-10-328-475-22	Sequence 22, Appl
451	7	1.7	615	11	US-09-997-601-194	Sequence 194, App	524	6	1.5	66	10	US-09-728-626-3840	Sequence 3840, App
452	7	1.7	615	11	US-09-990-443-194	Sequence 194, App	525	6	1.5	67	10	US-09-764-877-1873	Sequence 1873, App
453	7	1.7	615	11	US-09-991-854-194	Sequence 194, App	526	6	1.5	70	10	US-09-289-346A-3	Sequence 3, Appl

Tue Sep 9 08:55:30 2003

527	6	1.5	70	10	US-09-289-346A-14	Sequence 14, Appl	600	6	1.5	142	15	US-10-156-761-7740	Sequence 7740, Ap
528	6	1.5	70	10	US-09-796-692-2365	Sequence 2365, Ap	601	6	1.5	142	15	US-10-156-761-10346	Sequence 10346, A
529	6	1.5	70	14	US-10-001-835-159	Sequence 159, App	602	6	1.5	142	15	US-10-156-761-12014	Sequence 12014, A
530	6	1.5	70	15	US-10-040-862-2365	Sequence 2365, Ap	603	6	1.5	143	10	US-09-738-626-5904	Sequence 5904, Ap
531	6	1.5	71	11	US-09-764-891-4600	Sequence 4600, Ap	604	6	1.5	143	15	US-10-156-761-9839	Sequence 9839, Ap
532	6	1.5	72	15	US-10-106-698-6127	Sequence 6127, Ap	605	6	1.5	144	11	US-09-764-881-133	Sequence 133, App
533	6	1.5	73	10	US-09-738-626-5102	Sequence 5102, Ap	606	6	1.5	147	9	US-09-764-864-1025	Sequence 1025, Ap
534	6	1.5	73	15	US-10-083-357-944	Sequence 944, App	607	6	1.5	149	9	US-09-940-497-2	Sequence 2, Appl1
535	6	1.5	74	11	US-09-974-879-244	Sequence 244, App	608	6	1.5	149	11	US-09-764-891-5349	Sequence 5349, Ap
536	6	1.5	75	11	US-09-305-736-245	Sequence 245, App	609	6	1.5	149	14	US-10-004-717-3	Sequence 23, Appl
537	6	1.5	77	11	US-09-918-036-3	Sequence 3, Appl1	610	6	1.5	150	10	US-09-731-872-363	Sequence 363, App
538	6	1.5	78	9	US-09-864-761-37650	Sequence 37650, A	611	6	1.5	150	10	US-09-738-626-4598	Sequence 4598, Ap
539	6	1.5	79	9	US-09-864-761-47232	Sequence 47232, A	612	6	1.5	150	12	US-09-876-997-363	Sequence 363, App
540	6	1.5	80	11	US-09-776-724A-139	Sequence 139, App	613	6	1.5	152	10	US-09-738-626-5856	Sequence 5856, Ap
541	6	1.5	80	11	US-09-948-820-72	Sequence 72, Appl	614	6	1.5	153	12	US-10-017-161-1854	Sequence 1854, Ap
542	6	1.5	86	15	US-10-011-585A-154	Sequence 154, App	615	6	1.5	156	12	US-10-017-161-1906	Sequence 1906, Ap
543	6	1.5	87	10	US-09-738-626-3513	Sequence 3513, Ap	616	6	1.5	157	15	US-10-156-761-10551	Sequence 10551, A
544	6	1.5	87	11	US-09-938-806A-5	Sequence 5, Appl1	617	6	1.5	157	15	US-10-156-761-12728	Sequence 12728, A
545	6	1.5	87	15	US-10-072-783-6	Sequence 6, Appl1	618	6	1.5	159	15	US-10-156-761-13807	Sequence 13807, A
546	6	1.5	88	11	US-09-764-891-3010	Sequence 3010, Ap	619	6	1.5	161	12	US-10-017-161-1070	Sequence 1070, Ap
547	6	1.5	90	10	US-09-818-247-17	Sequence 17, Appl	620	6	1.5	161	12	US-10-238-075-1254	Sequence 1254, Ap
548	6	1.5	90	11	US-09-818-247-17	Sequence 25, App	621	6	1.5	161	15	US-10-156-761-11117	Sequence 11117, A
549	6	1.5	90	14	US-10-001-873-34	Sequence 34, Appl	622	6	1.5	162	10	US-09-738-626-4267	Sequence 4267, A
550	6	1.5	94	10	US-09-925-300-1157	Sequence 1157, App	623	6	1.5	163	12	US-10-017-161-1894	Sequence 1894, Ap
551	6	1.5	94	11	US-09-764-872-412	Sequence 412, App	624	6	1.5	164	15	US-10-156-761-13229	Sequence 13229, A
552	6	1.5	95	9	US-09-925-301-892	Sequence 892, App	625	6	1.5	167	12	US-10-017-161-1154	Sequence 1154, Ap
553	6	1.5	95	10	US-09-818-247-20	Sequence 20, Appl	626	6	1.5	168	15	US-10-156-761-8638	Sequence 8638, Ap
554	6	1.5	96	11	US-09-764-891-3006	Sequence 3006, Ap	627	6	1.5	169	10	US-09-738-626-4159	Sequence 4159, Ap
555	6	1.5	96	14	US-10-093-892-7	Sequence 7, Appl1	628	6	1.5	170	10	US-09-839-894-36	Sequence 839, Ap
556	6	1.5	96	14	US-10-093-892-9	Sequence 9, Appl1	629	6	1.5	173	15	US-10-156-761-7665	Sequence 7665, Ap
557	6	1.5	96	14	US-10-093-892-9	Sequence 37100, A	630	6	1.5	175	9	US-09-815-242-11637	Sequence 11637, A
558	6	1.5	101	9	US-09-864-761-37100	Sequence 37100, A	631	6	1.5	175	15	US-10-156-761-11876	Sequence 11876, A
559	6	1.5	102	12	US-09-969-748C-104	Sequence 20, Appl	632	6	1.5	175	15	US-10-156-761-9974	Sequence 9974, Ap
560	6	1.5	103	15	US-10-056-884-20	Sequence 5, Appl1	633	6	1.5	176	15	US-10-156-761-9187	Sequence 9187, A
561	6	1.5	105	14	US-10-093-892-5	Sequence 84, Appl	634	6	1.5	177	8	US-08-424-550B-49	Sequence 49, Appl
562	6	1.5	107	11	US-09-820-843A-84	Sequence 143, App	635	6	1.5	177	11	US-09-963-347B-7	Sequence 7, Appl1
563	6	1.5	108	15	US-10-078-030-143	Sequence 6169, Ap	636	6	1.5	177	11	US-09-902-517-18	Sequence 18, Appl
564	6	1.5	109	10	US-09-738-626-6169	Sequence 108, App	637	6	1.5	177	12	US-10-400-377-14	Sequence 14, Appl
565	6	1.5	112	11	US-09-948-820-108	Sequence 334, App	638	6	1.5	180	11	US-09-905-666A-75	Sequence 75, Appl
566	6	1.5	115	15	US-10-043-487-334	Sequence 8675, Ap	639	6	1.5	180	11	US-09-905-666A-76	Sequence 76, Appl
567	6	1.5	118	15	US-10-156-761-8675	Sequence 1172, Ap	640	6	1.5	180	11	US-09-905-666A-77	Sequence 77, Appl
568	6	1.5	119	10	US-09-764-868-1172	Sequence 14, Appl	641	6	1.5	180	11	US-09-905-666A-78	Sequence 78, Appl
569	6	1.5	121	10	US-09-775-932-14	Sequence 1641, App	642	6	1.5	180	11	US-09-905-666A-79	Sequence 79, Appl
570	6	1.5	121	15	US-10-156-761-13625	Sequence 13625, A	643	6	1.5	180	11	US-09-905-666A-80	Sequence 80, Appl
571	6	1.5	123	11	US-09-809-391-693	Sequence 620, App	644	6	1.5	180	11	US-09-905-666A-81	Sequence 81, Appl
572	6	1.5	125	9	US-09-867-550-620	Sequence 6343, Ap	645	6	1.5	180	11	US-09-905-666A-82	Sequence 82, Appl
573	6	1.5	125	10	US-09-738-626-6343	Sequence 20, Appl	646	6	1.5	180	11	US-09-905-666A-83	Sequence 83, Appl
574	6	1.5	126	9	US-09-934-586A-20	Sequence 8612, Ap	647	6	1.5	180	11	US-09-905-666A-84	Sequence 84, Appl
575	6	1.5	126	15	US-10-156-761-8612	Sequence 12966, A	648	6	1.5	180	11	US-09-905-666A-85	Sequence 85, Appl
576	6	1.5	127	9	US-09-815-242-12966	Sequence 12, Appl	649	6	1.5	180	11	US-09-905-666A-86	Sequence 86, Appl
577	6	1.5	128	10	US-09-775-932-12	Sequence 538, App	650	6	1.5	180	11	US-09-905-666A-87	Sequence 87, Appl
578	6	1.5	130	9	US-09-764-853-538	Sequence 171, App	651	6	1.5	180	11	US-09-905-666A-88	Sequence 88, Appl
579	6	1.5	130	15	US-10-091-438-171	Sequence 10984, A	652	6	1.5	180	11	US-09-905-666A-89	Sequence 89, Appl
580	6	1.5	131	15	US-10-156-761-10984	Sequence 89, Appl	653	6	1.5	180	11	US-09-905-666A-90	Sequence 90, Appl
581	6	1.5	132	9	US-09-789-561-89	Sequence 474, App	654	6	1.5	180	11	US-09-905-666A-91	Sequence 91, Appl
582	6	1.5	134	10	US-09-731-872-474	Sequence 474, App	655	6	1.5	180	11	US-09-905-666A-92	Sequence 92, Appl
583	6	1.5	134	10	US-09-731-872-474	Sequence 475, App	656	6	1.5	180	11	US-09-905-666A-93	Sequence 93, Appl
584	6	1.5	134	10	US-09-738-626-5042	Sequence 474, App	657	6	1.5	180	11	US-09-905-666A-94	Sequence 94, Appl
585	6	1.5	134	12	US-09-876-997-474	Sequence 475, App	658	6	1.5	180	11	US-09-905-666A-95	Sequence 95, Appl
586	6	1.5	134	12	US-09-876-997-475	Sequence 123, App	659	6	1.5	180	11	US-09-905-666A-96	Sequence 96, Appl
587	6	1.5	136	10	US-09-738-973-123	Sequence 123, App	660	6	1.5	180	11	US-09-905-666A-97	Sequence 97, Appl
588	6	1.5	136	10	US-09-854-133-123	Sequence 123, App	661	6	1.5	180	11	US-09-905-666A-98	Sequence 98, Appl
589	6	1.5	136	15	US-10-144-649A-123	Sequence 12625, A	662	6	1.5	180	11	US-09-905-666A-99	Sequence 99, Appl
590	6	1.5	137	15	US-10-156-761-12625	Sequence 210, App	663	6	1.5	180	11	US-09-905-666A-100	Sequence 100, App
591	6	1.5	138	11	US-09-468-147-210	Sequence 210, App	664	6	1.5	180	11	US-09-905-666A-101	Sequence 101, App
592	6	1.5	138	11	US-09-468-147-214	Sequence 248, App	665	6	1.5	180	11	US-09-905-666A-102	Sequence 102, App
593	6	1.5	138	11	US-09-468-147-244	Sequence 248, App	666	6	1.5	180	11	US-09-905-666A-103	Sequence 103, App
594	6	1.5	138	11	US-09-468-147-248	Sequence 252, App	667	6	1.5	180	11	US-09-905-666A-104	Sequence 104, App
595	6	1.5	138	11	US-09-468-147-252	Sequence 162, App	668	6	1.5	180	11	US-09-905-666A-105	Sequence 105, App
596	6	1.5	138	11	US-09-764-881-142	Sequence 289, App	669	6	1.5	180	11	US-09-905-666A-106	Sequence 106, App
597	6	1.5	140	9	US-09-764-887-289	Sequence 66, Appl	670	6	1.5	180	11	US-09-905-666A-107	Sequence 107, App
598	6	1.5	140	12	US-09-890-688-68	Sequence 289, App	671	6	1.5	180	11	US-09-905-666A-108	Sequence 108, App
599	6	1.5	140	15	US-10-073-961-289	Sequence 289, App	672	6	1.5	181	11	US-09-848-616-155	Sequence 155, App

819	6	1.5	204	15	US-10-184-622-40	Sequence 40, Appl	832	6	1.5	204	15	US-10-179-522-40	Sequence 40, Appl
820	6	1.5	204	15	US-10-184-628-40	Sequence 40, Appl	833	6	1.5	204	15	US-10-180-556-40	Sequence 40, Appl
821	6	1.5	204	15	US-10-184-629-40	Sequence 40, Appl	894	6	1.5	204	15	US-10-180-560-40	Sequence 40, Appl
822	6	1.5	204	15	US-10-184-630-40	Sequence 40, Appl	895	6	1.5	204	15	US-10-183-015-40	Sequence 40, Appl
823	6	1.5	204	15	US-10-184-631-40	Sequence 40, Appl	896	6	1.5	204	15	US-10-184-615-40	Sequence 40, Appl
824	6	1.5	204	15	US-10-184-632-40	Sequence 40, Appl	897	6	1.5	204	15	US-10-184-620-40	Sequence 40, Appl
825	6	1.5	204	15	US-10-184-636-40	Sequence 40, Appl	898	6	1.5	204	15	US-10-184-643-40	Sequence 40, Appl
826	6	1.5	204	15	US-10-184-640-40	Sequence 40, Appl	899	6	1.5	204	15	US-10-184-656-40	Sequence 40, Appl
827	6	1.5	204	15	US-10-184-650-40	Sequence 40, Appl	900	6	1.5	204	15	US-10-192-010-40	Sequence 40, Appl
828	6	1.5	204	15	US-10-184-651-40	Sequence 40, Appl	901	6	1.5	204	15	US-10-205-908-40	Sequence 40, Appl
829	6	1.5	204	15	US-10-187-588-40	Sequence 40, Appl	902	6	1.5	204	15	US-10-186-855-40	Sequence 40, Appl
830	6	1.5	204	15	US-10-187-597-40	Sequence 40, Appl	903	6	1.5	204	15	US-10-017-081A-36	Sequence 36, Appl
831	6	1.5	204	15	US-10-187-598-40	Sequence 40, Appl	904	6	1.5	204	15	US-10-184-619-40	Sequence 40, Appl
832	6	1.5	204	15	US-10-187-600-40	Sequence 40, Appl	905	6	1.5	204	15	US-10-187-599-40	Sequence 40, Appl
833	6	1.5	204	15	US-10-187-601-40	Sequence 40, Appl	906	6	1.5	204	15	US-10-187-750-40	Sequence 40, Appl
834	6	1.5	204	15	US-10-187-602-40	Sequence 40, Appl	907	6	1.5	204	15	US-10-188-780-40	Sequence 40, Appl
835	6	1.5	204	15	US-10-187-603-40	Sequence 40, Appl	908	6	1.5	204	15	US-10-192-015-40	Sequence 40, Appl
836	6	1.5	204	15	US-10-187-741-40	Sequence 40, Appl	909	6	1.5	204	15	US-10-194-394-40	Sequence 40, Appl
837	6	1.5	204	15	US-10-187-743-40	Sequence 40, Appl	910	6	1.5	204	15	US-10-194-425-40	Sequence 40, Appl
838	6	1.5	204	15	US-10-187-746-40	Sequence 40, Appl	911	6	1.5	204	15	US-10-194-485-40	Sequence 40, Appl
839	6	1.5	204	15	US-10-187-747-40	Sequence 40, Appl	912	6	1.5	204	15	US-10-195-885-40	Sequence 40, Appl
840	6	1.5	204	15	US-10-187-751-40	Sequence 40, Appl	913	6	1.5	204	15	US-10-195-899-40	Sequence 40, Appl
841	6	1.5	204	15	US-10-187-753-40	Sequence 40, Appl	914	6	1.5	204	15	US-10-196-748-40	Sequence 40, Appl
842	6	1.5	204	15	US-10-187-754-40	Sequence 40, Appl	915	6	1.5	204	15	US-10-196-750-40	Sequence 40, Appl
843	6	1.5	204	15	US-10-187-757-40	Sequence 40, Appl	916	6	1.5	204	15	US-10-197-699-40	Sequence 40, Appl
844	6	1.5	204	15	US-10-187-884-40	Sequence 40, Appl	917	6	1.5	204	15	US-10-197-700-40	Sequence 40, Appl
845	6	1.5	204	15	US-10-188-767-40	Sequence 40, Appl	918	6	1.5	204	15	US-10-197-705-40	Sequence 40, Appl
846	6	1.5	204	15	US-10-188-769-40	Sequence 40, Appl	919	6	1.5	204	15	US-10-197-708-40	Sequence 40, Appl
847	6	1.5	204	15	US-10-188-770-40	Sequence 40, Appl	920	6	1.5	204	15	US-10-198-764-40	Sequence 40, Appl
848	6	1.5	204	15	US-10-188-773-40	Sequence 40, Appl	921	6	1.5	204	15	US-10-198-765-40	Sequence 40, Appl
849	6	1.5	204	15	US-10-188-781-40	Sequence 40, Appl	922	6	1.5	204	15	US-10-198-768-40	Sequence 40, Appl
850	6	1.5	204	15	US-10-194-361-40	Sequence 40, Appl	923	6	1.5	204	15	US-10-198-769-40	Sequence 40, Appl
851	6	1.5	204	15	US-10-194-423-40	Sequence 40, Appl	924	6	1.5	204	15	US-10-199-317-40	Sequence 40, Appl
852	6	1.5	204	15	US-10-195-897-40	Sequence 40, Appl	925	6	1.5	204	15	US-10-199-306-40	Sequence 40, Appl
853	6	1.5	204	15	US-10-195-901-40	Sequence 40, Appl	926	6	1.5	204	15	US-10-199-310-40	Sequence 40, Appl
854	6	1.5	204	15	US-10-195-902-40	Sequence 40, Appl	927	6	1.5	204	15	US-10-199-311-40	Sequence 40, Appl
855	6	1.5	204	15	US-10-196-743-40	Sequence 40, Appl	928	6	1.5	204	15	US-10-199-314-40	Sequence 40, Appl
856	6	1.5	204	15	US-10-196-760-40	Sequence 40, Appl	929	6	1.5	204	15	US-10-199-317-40	Sequence 40, Appl
857	6	1.5	204	15	US-10-173-708-40	Sequence 40, Appl	930	6	1.5	204	15	US-10-199-665-40	Sequence 40, Appl
858	6	1.5	204	15	US-10-176-479-40	Sequence 40, Appl	931	6	1.5	204	15	US-10-199-666-40	Sequence 40, Appl
859	6	1.5	204	15	US-10-176-748-40	Sequence 40, Appl	932	6	1.5	204	15	US-10-201-534-40	Sequence 40, Appl
860	6	1.5	204	15	US-10-176-916-40	Sequence 40, Appl	933	6	1.5	204	15	US-10-201-770-40	Sequence 40, Appl
861	6	1.5	204	15	US-10-179-507-40	Sequence 40, Appl	934	6	1.5	204	15	US-10-201-885-40	Sequence 40, Appl
862	6	1.5	204	15	US-10-179-519-40	Sequence 40, Appl	935	6	1.5	204	15	US-10-202-936-40	Sequence 40, Appl
863	6	1.5	204	15	US-10-179-525-40	Sequence 40, Appl	936	6	1.5	204	15	US-10-202-938-40	Sequence 40, Appl
864	6	1.5	204	15	US-10-180-540-40	Sequence 40, Appl	937	6	1.5	204	15	US-10-202-463-40	Sequence 40, Appl
865	6	1.5	204	15	US-10-180-545-40	Sequence 40, Appl	938	6	1.5	204	15	US-10-202-470-40	Sequence 40, Appl
866	6	1.5	204	15	US-10-183-006-40	Sequence 40, Appl	939	6	1.5	204	15	US-10-202-476-40	Sequence 40, Appl
867	6	1.5	204	15	US-10-183-008-40	Sequence 40, Appl	940	6	1.5	204	15	US-10-202-934-40	Sequence 40, Appl
868	6	1.5	204	15	US-10-183-017-40	Sequence 40, Appl	941	6	1.5	204	15	US-10-202-935-40	Sequence 40, Appl
869	6	1.5	204	15	US-10-183-019-40	Sequence 40, Appl	942	6	1.5	204	15	US-10-202-936-40	Sequence 40, Appl
870	6	1.5	204	15	US-10-184-618-40	Sequence 40, Appl	943	6	1.5	204	15	US-10-202-939-40	Sequence 40, Appl
871	6	1.5	204	15	US-10-184-623-40	Sequence 40, Appl	944	6	1.5	204	15	US-10-205-504-40	Sequence 40, Appl
872	6	1.5	204	15	US-10-184-625-40	Sequence 40, Appl	945	6	1.5	204	15	US-10-205-509-40	Sequence 40, Appl
873	6	1.5	204	15	US-10-184-626-40	Sequence 40, Appl	946	6	1.5	204	15	US-10-205-895-40	Sequence 40, Appl
874	6	1.5	204	15	US-10-184-627-40	Sequence 40, Appl	947	6	1.5	204	15	US-10-205-899-40	Sequence 40, Appl
875	6	1.5	204	15	US-10-184-645-40	Sequence 40, Appl	948	6	1.5	204	15	US-10-205-900-40	Sequence 40, Appl
876	6	1.5	204	15	US-10-184-654-40	Sequence 40, Appl	949	6	1.5	204	15	US-10-205-909-40	Sequence 40, Appl
877	6	1.5	204	15	US-10-184-655-40	Sequence 40, Appl	950	6	1.5	204	15	US-10-199-890-40	Sequence 40, Appl
878	6	1.5	204	15	US-10-188-774-40	Sequence 40, Appl	951	6	1.5	204	15	US-10-184-602-40	Sequence 40, Appl
879	6	1.5	204	15	US-10-188-775-40	Sequence 40, Appl	952	6	1.5	204	15	US-10-184-621-40	Sequence 40, Appl
880	6	1.5	204	15	US-10-194-462-40	Sequence 40, Appl	953	6	1.5	204	15	US-10-184-638-40	Sequence 40, Appl
881	6	1.5	204	15	US-10-196-745-40	Sequence 40, Appl	954	6	1.5	204	15	US-10-187-753-40	Sequence 40, Appl
882	6	1.5	204	15	US-10-196-762-40	Sequence 40, Appl	955	6	1.5	204	15	US-10-187-887-40	Sequence 40, Appl
883	6	1.5	204	15	US-10-197-695-40	Sequence 40, Appl	956	6	1.5	204	15	US-10-194-461-40	Sequence 40, Appl
884	6	1.5	204	15	US-10-195-894-40	Sequence 40, Appl	957	6	1.5	204	15	US-10-195-892-40	Sequence 40, Appl
885	6	1.5	204	15	US-10-176-84-40	Sequence 40, Appl	958	6	1.5	204	15	US-10-196-751-40	Sequence 40, Appl
886	6	1.5	204	15	US-10-176-753-40	Sequence 40, Appl	959	6	1.5	204	15	US-10-197-694-40	Sequence 40, Appl
887	6	1.5	204	15	US-10-176-917-40	Sequence 40, Appl	960	6	1.5	204	15	US-10-197-697-40	Sequence 40, Appl
888	6	1.5	204	15	US-10-176-982-40	Sequence 40, Appl	961	6	1.5	204	15	US-10-197-707-40	Sequence 40, Appl
889	6	1.5	204	15	US-10-179-506-40	Sequence 40, Appl	962	6	1.5	204	15	US-10-199-303-40	Sequence 40, Appl
890	6	1.5	204	15	US-10-179-513-40	Sequence 40, Appl	963	6	1.5	204	15	US-10-199-318-40	Sequence 40, Appl
891	6	1.5	204	15	US-10-179-514-40	Sequence 40, Appl	964	6	1.5	204	15	US-10-199-458-40	Sequence 40, Appl

965 6 1.5 204 15 US-10-199-462-40 Sequence 40, Appl
966 6 1.5 204 15 US-10-201-324-40 Sequence 40, Appl
967 6 1.5 204 15 US-10-201-328-40 Sequence 40, Appl
968 6 1.5 204 15 US-10-201-527-40 Sequence 40, Appl
969 6 1.5 204 15 US-10-201-528-40 Sequence 40, Appl
970 6 1.5 204 15 US-10-201-529-40 Sequence 40, Appl
971 6 1.5 204 15 US-10-201-530-40 Sequence 40, Appl
972 6 1.5 204 15 US-10-202-408-40 Sequence 40, Appl
973 6 1.5 204 15 US-10-202-409-40 Sequence 40, Appl
974 6 1.5 204 15 US-10-202-411-40 Sequence 40, Appl
975 6 1.5 204 15 US-10-202-472-40 Sequence 40, Appl
976 6 1.5 204 15 US-10-205-502-40 Sequence 40, Appl
977 6 1.5 204 15 US-10-205-507-40 Sequence 40, Appl
978 6 1.5 204 15 US-10-205-511-40 Sequence 40, Appl
979 6 1.5 204 15 US-10-205-902-40 Sequence 40, Appl
980 6 1.5 204 15 US-10-205-907-40 Sequence 40, Appl
981 6 1.5 204 15 US-10-167-749-36 Sequence 36, Appl
982 6 1.5 204 15 US-10-194-456-40 Sequence 40, Appl
983 6 1.5 204 15 US-10-196-758-40 Sequence 40, Appl
984 6 1.5 204 15 US-10-198-770-40 Sequence 40, Appl
985 6 1.5 204 15 US-10-198-308-40 Sequence 40, Appl
986 6 1.5 204 15 US-10-200-617-40 Sequence 40, Appl
987 6 1.5 204 15 US-10-205-893-40 Sequence 40, Appl
988 6 1.5 204 15 US-10-205-897-40 Sequence 40, Appl
989 6 1.5 204 15 US-10-196-754-40 Sequence 40, Appl
990 6 1.5 204 15 US-10-176-847-8 Sequence 8, Appl
991 6 1.5 204 15 US-10-012-921A-36 Sequence 36, Appl
992 6 1.5 204 15 US-10-174-571-40 Sequence 40, Appl
993 6 1.5 204 15 US-10-176-746-40 Sequence 40, Appl
994 6 1.5 204 15 US-10-176-923-40 Sequence 40, Appl
995 6 1.5 204 15 US-10-183-011-40 Sequence 40, Appl
996 6 1.5 204 15 US-10-184-633-40 Sequence 40, Appl
997 6 1.5 204 15 US-10-184-639-40 Sequence 40, Appl
998 6 1.5 204 15 US-10-187-742-40 Sequence 40, Appl
999 6 1.5 204 15 US-10-187-748-40 Sequence 40, Appl
1000 6 1.5 204 15 US-10-206-907-40 Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-10-012-542-398
Sequence 398, Application US/10012542
Publication No. US20030044851A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 94 Human Secreted Proteins
FILE REFERENCE: P2029P1
CURRENT APPLICATION NUMBER: US/10/012,542
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
NUMBER OF SEQ ID NOS: 532
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 398
LENGTH: 124
TYPE: PRT
ORGANISM: Homo sapiens
US-10-012-542-398

Query Match 2.4%; Score 10; DB 15; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALAAVA 19
|||||
Db 104 RAALAAVA 113

RESULT 2
US-09-890-688-154
Sequence 154, Application US/09890688
Publication No. US20030144475A1
GENERAL INFORMATION:
APPLICANT: Seiichi KATO
APPLICANT: Chikashi EGUCHI
TITLE OF INVENTION: Human Proteins and cDNAs thereof
FILE REFERENCE: 2001-1102A/MMC/00653
CURRENT APPLICATION NUMBER: US/09/890,688
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: JP 11-346863
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: JP 11-34684
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: JP 2000-31062
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: JP 2000-34091
PRIOR FILING DATE: 2000-02-10
PRIOR APPLICATION NUMBER: JP 2000-34090
PRIOR FILING DATE: 2000-02-10
PRIOR APPLICATION NUMBER: JP 2000-35829
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: JP 2000-35899
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: JP 2000-71161
PRIOR FILING DATE: 2000-03-14
PRIOR APPLICATION NUMBER: JP 2000-160851
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 160
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 154
LENGTH: 222
TYPE: PRT
ORGANISM: Homo sapiens
US-09-890-688-154

Query Match 2.4%; Score 10; DB 12; Length 222;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALAAVA 19
|||||
Db 202 RAALAAVA 211

RESULT 3
US-10-156-761-10034
Sequence 10034, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10034
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10034

Query Match
Best Local Similarity 1.9%; Score 8; DB 15; Length 102;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 AAAVALVL 22
|||
Db 72 AAAVALVL 79

RESULT 4
US-10-080-170-577
; Sequence 577, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 577
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-577

Query Match
Best Local Similarity 1.9%; Score 8; DB 16; Length 181;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAAV 18
|||
Db 16 AAALAAV 23

RESULT 5
US-10-012-542-397
; Sequence 397, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029p1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 397
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-542-397

Query Match
Best Local Similarity 1.9%; Score 8; DB 15; Length 194;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAVA 19
|||
Db 176 AALAAVA 183

RESULT 6
US-09-813-555-1
; Sequence 1, Application US/09813555
; Patent No. US20010034455A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (Mecap2) Inhibitors
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-813-555-1

Query Match
Best Local Similarity 1.9%; Score 8; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||
Db 235 NAGDTTVL 242

RESULT 7
US-09-813-555-2
; Sequence 2, Application US/09813555
; Patent No. US20010034455A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (Mecap2) Inhibitors
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-813-555-2

Query Match
Best Local Similarity 1.9%; Score 8; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTTL 210
|||||
Db 235 NAGDTTTL 242

RESULT 8

US-09-813-555-3
; Sequence 3, Application US/09813555
; Patent No. US20010034455A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-813-555-3

Query Match 1.9%; Score 8; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTTL 210
|||||
Db 235 NAGDTTTL 242

RESULT 9

US-09-943-123-6
; Sequence 6, Application US/09943123
; Publication No. US20020182701A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human dmymetap2

FEATURE:
NAME/KEY: SITE
LOCATION: (219)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (231)
OTHER INFORMATION: May be any amino acid, except His
NAME/KEY: SITE
LOCATION: (251)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (262)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (328)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (331)
OTHER INFORMATION: May be any naturally occurring amino acid

NAME/KEY: SITE
LOCATION: (338)..(339)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (364)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (444)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (447)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (459)
OTHER INFORMATION: May be any naturally occurring amino acid
US-09-943-123-6

Query Match 1.9%; Score 8; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTTL 210
|||||
Db 235 NAGDTTTL 242

RESULT 10

US-09-943-123-7
; Sequence 7, Application US/09943123
; Publication No. US20020182701A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mouse Metap2

FEATURE:
NAME/KEY: SITE
LOCATION: (219)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (231)
OTHER INFORMATION: May be any amino acid, except His
NAME/KEY: SITE
LOCATION: (251)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (262)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (328)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (331)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (338)..(339)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (364)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (444)
OTHER INFORMATION: May be any naturally occurring amino acid

NAME/KEY: SITE
LOCATION: (447)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (459)
OTHER INFORMATION: May be any naturally occurring amino acid
US-09-943-123-7

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 478;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 11
US-09-943-123-12
Sequence 12, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:

APPLICANT: CHANG, Y-H
APPLICANT: VETRO, J.A.
APPLICANT: MICKA, W.S.
TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
FILE REFERENCE: 16153-8007
CURRENT APPLICATION NUMBER: US/09/943,123
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEO ID NO 12
LENGTH: 478
TYPE: PRT
ORGANISM: Human Metap2
US-09-943-123-12

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 478;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 12
US-09-943-123-13
Sequence 13, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:

APPLICANT: CHANG, Y-H
APPLICANT: VETRO, J.A.
APPLICANT: MICKA, W.S.
TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
FILE REFERENCE: 16153-8007
CURRENT APPLICATION NUMBER: US/09/943,123
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEO ID NO 13
LENGTH: 478
TYPE: PRT
ORGANISM: Mouse Metap2
US-09-943-123-13

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 478;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210

Db 235 NAGDTTVL 242

RESULT 13
US-09-943-123-16
Sequence 16, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:
APPLICANT: CHANG, Y-H
APPLICANT: VETRO, J.A.
APPLICANT: MICKA, W.S.
TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
FILE REFERENCE: 16153-8007
CURRENT APPLICATION NUMBER: US/09/943,123
CURRENT FILING DATE: 2001-08-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEO ID NO 16
LENGTH: 480
TYPE: PRT
ORGANISM: Rat dnvMetap2
FEATURE:
NAME/KEY: SITE
LOCATION: (219)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (231)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (251)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (262)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (328)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (331)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (338)..
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (364)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (444)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (447)
OTHER INFORMATION: May be any naturally occurring amino acid
NAME/KEY: SITE
LOCATION: (459)
OTHER INFORMATION: May be any naturally occurring amino acid
US-09-943-123-16

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 480;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 14
US-09-943-123-17
Sequence 17, Application US/09943123
Publication No. US20020182701A1
GENERAL INFORMATION:

```

; APPLICANT: CHANG, Y-H
; APPLICANT: VETRO, J.A.
; APPLICANT: MICKA, W.S.
; TITLE OF INVENTION: Dominant Negative Variants of Methionine Aminopeptidase
; FILE REFERENCE: 16153-8007
; CURRENT APPLICATION NUMBER: US/09/943,123
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 17
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Rat MetAP2
US-09-943-123-17

Query Match
Best Local Similarity 1.9%; Score 8; DB 10; Length 480;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 15
US-10-106-698-4765
; Sequence 4765, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA0050P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4765
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4765

Query Match
Best Local Similarity 1.9%; Score 8; DB 15; Length 500;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NAGDTTVL 210
Db 257 NAGDTTVL 264

RESULT 16
US-09-815-242-10147
; Sequence 10147, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
```

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10147
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10147

Query Match
Best Local Similarity 1.9%; Score 8; DB 9; Length 891;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 RAALALAA 17
Db 36 RAALALAA 43

RESULT 17
US-09-815-242-13843
; Sequence 13843, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13843
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
```

OTHER INFORMATION: EXPRESSED IN BT47

PRIOR APPLICATION NUMBER: EARLIER APPLICA

! PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18

```

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 347
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-347

```

```

Query Match      1.7%; Score 7; DB 11; Length 24;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      111 NLESARA 117
      |||||
Db      8 NLESARA 14

```

```

RESULT 20
US-09-864-761-44448
; Sequence 44448, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44448
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005039.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86
; OTHER INFORMATION: EST_HUMAN HIT: AW512200.1, EVALU8 1.30e-01
US-09-864-761-44448

```

```

Query Match      1.7%; Score 7; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      66 RLESIRT 72
      |||||
Db      2 RLESIRT 8

```

```

RESULT 21
US-09-764-877-1958
; Sequence 1958, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1958
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (117)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (63)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1958

```

```

Query Match      1.7%; Score 7; DB 10; Length 65;

```

Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 232 NIMMGL 298
DB 20 NIMMGL 26

RESULT 22

US-10-156-761-13107
; Sequence 13107, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13107
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13107

Query Match 1.7%; Score 7; DB 15; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAA 17
DB 55 AAALAA 61

RESULT 23

US-10-007-267-2
; Sequence 2, Application US/10007267
; Publication No. US20020127682A1
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/007,267
; FILING DATE: 03-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,412
; FILING DATE: 15-Jun-1999
; APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-007-267-2

Query Match 1.7%; Score 7; DB 14; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAA 17
DB 13 AAALAA 19

RESULT 24

US-09-956-425-4
; Sequence 4, Application US/09956425
; Patent No. US20020045192A1
; GENERAL INFORMATION:
; APPLICANT: Kriwacki, Richard
; APPLICANT: Bochner, Brian
; APPLICANT: Lewis, William
; TITLE OF INVENTION: Aft and Hdm2 Interaction Domains and Method of Use Thereof
; FILE REFERENCE: 1340/1/035
; CURRENT APPLICATION NUMBER: US/09/956,425
; CURRENT FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent version 3.1
; SEQ ID NO 4
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-956-425-4

Query Match 1.7%; Score 7; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALV 22
DB 41 AAVALV 47

RESULT 25

US-09-867-550-150
; Sequence 150, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrtan, Fuad
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30


```
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-150
```

```
Query Match          1.7%; Score 7; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      12 AALAAV 18
        |||||
Db       53 AALAAV 59
```

```
RESULT 26
US-09-925-300-1688
; Sequence 1688, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1688
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1688
```

```
Query Match          1.7%; Score 7; DB 10; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      10 RAALAA 16
        |||||
Db       21 RAALAA 27
```

```
RESULT 27
US-10-156-761-10928
; Sequence 10928, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10928
; LENGTH: 175
; TYPE: PRT
```

```
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10928
```

```
Query Match          1.7%; Score 7; DB 15; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      392 AAPQSG 398
        |||||
Db       79 AAPQSG 85
```

```
RESULT 28
US-10-156-761-10784
; Sequence 10784, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10784
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10784
```

```
Query Match          1.7%; Score 7; DB 15; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      11 AAALAA 17
        |||||
Db       29 AAALAA 35
```

```
RESULT 29
US-09-764-853-558
; Sequence 558, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1206
; CURRENT APPLICATION NUMBER: US/09/764,853
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 558
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-853-558
```

```
Query Match          1.7%; Score 7; DB 9; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      41 EAPAPV 47
        |||||
```

Db 138 EAPAPV 144

RESULT 30

US-09-738-626-6652
 ; Sequence 6652, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738, 626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 6652
 ; LENGTH: 236
 ; TYPE: PR1
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6652

Query Match 1.7%; Score 7; DB 10; Length 236;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 366 VEGISIA 372
 Db 65 VEGISIA 71

RESULT 31

US-10-156-761-13873
 ; Sequence 13873, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272667
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 13873
 ; LENGTH: 248
 ; TYPE: PR1
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-13873

Query Match 1.7%; Score 7; DB 15; Length 248;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 233 LRQIAE 239
 Db 119 LRQIAE 125

RESULT 32

US-10-317-806-4
 ; Sequence 4, Application US/10317806
 ; Publication No. US20030113785A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Zayed, Adel
 ; APPLICANT: Ascenzi, Robert
 ; APPLICANT: Boyes, Douglas
 ; APPLICANT: Mulpur, Rao
 ; APPLICANT: Hoffman, Neil
 ; APPLICANT: Kjemtrup, Susanne
 ; APPLICANT: Davis, Keith
 ; APPLICANT: Phillips, Kenneth
 ; APPLICANT: Moesener, Jeffrey
 ; APPLICANT: Hamilton, Carol
 ; APPLICANT: Goriach, Jörn
 ; TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF HERBICIDES AND THE MODULATION O
 ; FILE REFERENCE: Docket # 2126US
 ; CURRENT APPLICATION NUMBER: US/10/317, 806
 ; CURRENT FILING DATE: 2002-12-12
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 253
 ; TYPE: PR1
 ; ORGANISM: Arabidopsis thaliana
 US-10-317-806-4

Query Match 1.7%; Score 7; DB 15; Length 253;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 403 SEAKTAS 409
 Db 180 SEAKTAS 186

RESULT 33

US-10-156-761-14170
 ; Sequence 14170, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 14170
 ; LENGTH: 262
 ; TYPE: PR1
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-14170

Query Match 1.7%; Score 7; DB 15; Length 262;

Query Match 1.7%; Score 7; DB 15; Length 262;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAA 16
 |||||
 216 RAALAA 222

RESULT 34
 US-10-342-224-72
 ; Sequence 72, Application US/10342224
 ; Publication No. US20030162294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nathalie Verbruggen
 ; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress
 ; FILE REFERENCE: CN-012US
 ; CURRENT APPLICATION NUMBER: US/10/342,224
 ; CURRENT FILING DATE: 2003-01-13
 ; PRIOR APPLICATION NUMBER: US/09/762,154
 ; PRIOR FILING DATE: 2002-02-02
 ; PRIOR APPLICATION NUMBER: EP 98202634.6
 ; PRIOR FILING DATE: 1998-08-04
 ; NUMBER OF SEQ ID NOS: 123
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 72
 ; LENGTH: 271
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; US-10-342-224-72

Query Match 1.7%; Score 7; DB 12; Length 271;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 245 ADGVIAV 251
 |||||
 136 ADGVIAV 142

RESULT 35
 US-09-815-242-10842
 ; Sequence 10842, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10842
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Enterococcus faecalis
 ; US-09-815-242-10842

Query Match 1.7%; Score 7; DB 9; Length 275;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 AVTRGAK 325
 |||||
 70 AVTRGAK 76

RESULT 36
 US-10-080-170-328
 ; Sequence 328, Application US/10080170
 ; Publication No. US20030129601A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cole, S.T.
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 ; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 ; FILE REFERENCE: 03495,0218
 ; CURRENT APPLICATION NUMBER: US/10/080,170
 ; CURRENT FILING DATE: 2002-06-10
 ; PRIOR APPLICATION NUMBER: 60/270,123
 ; PRIOR FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 652
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 328
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium leprae
 ; US-10-080-170-328

Query Match 1.7%; Score 7; DB 16; Length 275;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 QATLAKA 129
 |||||
 58 QATLAKA 64

RESULT 37
 US-10-156-761-14182
 ; Sequence 14182, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 14182
 ; LENGTH: 288
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-14182

Query Match 1.7%; Score 7; DB 15; Length 288;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALAAA 17
 DB 65 AALAAA 71

RESULT 38

US-10-156-761-8083
 ; Sequence 8083, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 8083
 ; LENGTH: 300
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-8083

Query Match 1.7%; Score 7; DB 15; Length 300;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 AGITGAK 378
 DB 259 AGITGAK 265

RESULT 39

US-10-156-761-13108
 ; Sequence 13108, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 13108
 ; LENGTH: 301
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-13108

Query Match 1.7%; Score 7; DB 15; Length 301;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 AADGVIA 250
 DB 121 AADGVIA 127

RESULT 40

US-09-533-029-28
 ; Sequence 28, Application US/09533029
 ; Publication No. US20030046723A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Heard, Jacqueline
 ; APPLICANT: Brown, Pierre
 ; APPLICANT: Riechmann, Jose-Luis
 ; APPLICANT: Keddie, James
 ; APPLICANT: Pineda, Omaira
 ; APPLICANT: Adam, Luc
 ; APPLICANT: Samaha, Raymond
 ; APPLICANT: Zhang, James
 ; APPLICANT: Yu, Guo-Liang
 ; APPLICANT: Ratcliffe, Oliver
 ; APPLICANT: Pilgrim, Marsha
 ; APPLICANT: Jiang, Cai-Zhong
 ; APPLICANT: Reuber, Lynne
 ; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
 ; FILE REFERENCE: MBI-010
 ; CURRENT APPLICATION NUMBER: US/09/533,029
 ; CURRENT FILING DATE: 2000-03-22
 ; EARLIER APPLICATION NUMBER: 60/125,814
 ; EARLIER FILING DATE: 1999-03-23
 ; NUMBER OF SEQ ID NOS: 121
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 302
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; OTHER INFORMATION: G291
 ; US-09-533-029-28

Query Match 1.7%; Score 7; DB 11; Length 302;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAV 18
 DB 284 AALAAV 290

Search completed: September 8, 2003, 14:15:02
 Job time: 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:06:52 ; Search time 41 Seconds
(without alignments)
2593.114 Million cell updates/sec

Title: US-09-889-756a-2

Perfect score: 412
Sequence: 1 MAFVAFKAMRAAALAAVAL.....AAPGSGVOTASEAKTASAE 412

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPRMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	41.5	412	16	Q9150 neisseria m
2	133	32.3	412	16	Q9156 neisseria m
3	100	24.3	271	2	Q51007 neisseria g
4	11	2.7	373	16	Q8X7E1 escherichia
5	11	2.7	374	2	Q8GC84 enterobacte
6	11	2.7	384	16	Q8CW42 escherichia
7	11	2.7	388	16	Q9AA03 caulobacter
8	10	2.4	124	4	Q969P9 homo sapien
9	10	2.4	216	4	Q81VM3 homo sapien
10	10	2.4	218	4	Q9BSM6 homo sapien
11	10	2.4	222	4	Q8TE51 homo sapien
12	10	2.4	222	4	Q8TAE8 homo sapien
13	9	2.2	444	16	Q8ZCW1 yersinia pe
14	8	1.9	107	16	Q9UQ07 neisseria m
15	8	1.9	125	10	Q94D46 oryza sativ
16	8	1.9	129	16	Q321B4 bacillus su

17	8	1.9	156	2	Q9RPU1 salmonella
18	8	1.9	162	10	Q8S0K1 oryza sativ
19	8	1.9	171	16	Q8XCR9 escherichia
20	8	1.9	183	16	Q8XC98 escherichia
21	8	1.9	183	17	Q9YAH2 aeropyrum p
22	8	1.9	206	16	Q8P176 xanthomonas
23	8	1.9	206	16	Q9KZV1 streptomyce
24	8	1.9	214	16	P95115 mycobacteri
25	8	1.9	222	6	Q8SEP7 cercopithec
26	8	1.9	238	2	Q9Z5W3 pseudomonas
27	8	1.9	253	4	Q96BA3 homo sapien
28	8	1.9	290	16	Q9S295 streptomyce
29	8	1.9	296	16	Q926G6 rhizobium m
30	8	1.9	306	16	Q9CH3 lactococcus
31	8	1.9	326	10	Q9SNL7 oryza sativ
32	8	1.9	348	16	Q67350 aquifex aeo
33	8	1.9	390	10	Q8GRS3 oryza sativ
34	8	1.9	394	2	Q9F241 xanthomonas
35	8	1.9	399	16	Q8P7C8 xanthomonas
36	8	1.9	408	16	Q8ZL01 salmonella
37	8	1.9	408	16	Q8Z2W0 salmonella
38	8	1.9	416	16	Q8XBZ9 escherichia
39	8	1.9	416	16	Q8FBW4 escherichia
40	8	1.9	442	16	Q8G7V0 biidobacte
41	8	1.9	444	5	Q9U2V9 caenorhabdi
42	8	1.9	451	5	Q93704 caenorhabdi
43	8	1.9	455	4	Q8NB11 homo sapien
44	8	1.9	559	16	Q8FTL2 corynebacte
45	8	1.9	654	16	Q8P9J8 xanthomonas
46	8	1.9	716	16	Q8ZUJ5 salmonella
47	8	1.9	748	16	Q8Z0W3 salmonella
48	8	1.9	748	16	Q9Z577 streptomyce
49	8	1.9	761	16	Q9KX15 streptomyce
50	8	1.9	825	5	Q8T3U6 drosophila
51	8	1.9	864	16	Q9RUB9 streptomyce
52	8	1.9	866	16	Q8EP55 shewanella
53	8	1.9	871	16	Q8ZG92 yersinia pe
54	8	1.9	878	2	P74880 salmonella
55	8	1.9	891	16	Q8ZED0 caenorhabdi
56	8	1.9	891	16	Q8FHX2 escherichia
57	8	1.9	892	16	Q8D0K2 yersinia pe
58	8	1.9	892	16	Q8ZPA5 salmonella
59	8	1.9	892	16	Q8Z7F1 salmonella
60	8	1.9	968	5	Q9VCB2 drosophila
61	8	1.9	979	10	Q23096 arabidopsis
62	8	1.9	984	2	Q9XCD4 thermomom
63	8	1.9	1121	5	P90630 acanthamoeb
64	8	1.9	1291	5	Q77261 drosophila
65	8	1.9	1300	5	Q9W5E0 drosophila
66	8	1.9	1487	5	Q8MPV5 caenorhabdi
67	8	1.9	1558	5	Q81770 caenorhabdi
68	8	1.9	2167	5	Q76840 caenorhabdi
69	8	1.9	2296	4	Q75042 homo sapien
70	8	1.7	37	2	Q81LD9 macaca faec
71	7	1.7	55	6	Q8S096 macaca faec
72	7	1.7	55	6	Q8S090 pan paniscu
73	7	1.7	55	6	Q8SCA2 cercopithec
74	7	1.7	55	6	Q8SQ99 gorilla gor
75	7	1.7	55	6	Q8SQ97 macaca arc
76	7	1.7	55	6	Q8SCQ3 cercopithec
77	7	1.7	63	10	Q91159 oryza sativ
78	7	1.7	72	2	Q8VB76 white spot
79	7	1.7	72	2	Q8RS17 uncultured
80	7	1.7	78	2	P95455 pseudomonas
81	7	1.7	80	10	Q24280 prunus pers
82	7	1.7	80	10	Q43750 capsicum an
83	7	1.7	81	2	Q9APP7 uncultured
84	7	1.7	87	4	Q9BV83 homo sapien
85	7	1.7	91	2	Q007301 pseudomonas
86	7	1.7	91	2	Q93M27 escherichia
87	7	1.7	91	15	Q8Q3B9 human immun
88	7	1.7	91	16	Q93511 salmonella
89	7	1.7	93	4	Q9B236 homo sapien

90	7	1.7	94	16	Q8P111	Q8P111 xanthomonas	163	7	1.7	203	16	Q9RK21	Q9RK21 streptomyc
91	7	1.7	96	4	Q9B237	Q9B237 homo sapien	164	7	1.7	204	2	Q8G909	Q8G909 helicobacte
92	7	1.7	97	4	Q9X4N8	Q9X4N8 pseudomonas	165	7	1.7	206	16	Q9X820	Q9X820 streptomyc
93	7	1.7	97	5	P91216	P91216 caenorhabdi	166	7	1.7	207	4	Q9BVM2	Q9BVM2 homo sapien
94	7	1.7	97	15	Q8Q3CO	Q8Q3CO human immun	167	7	1.7	208	6	Q29468	Q29468 canis famli
95	7	1.7	97	16	Q8K616	Q8K616 streptococc	168	7	1.7	208	10	Q94F04	Q94F04 lycopersico
96	7	1.7	100	11	Q8C808	Q8C808 mus musculu	169	7	1.7	209	16	Q9PLV8	Q9PLV8 campylobact
97	7	1.7	103	5	Q966Y0	Q966Y0 schistocerc	170	7	1.7	209	16	Q9A577	Q9A577 caulobacter
98	7	1.7	103	5	Q9SPM5	Q9SPM5 schistocerc	171	7	1.7	211	2	Q931I8	Q931I8 thibacilli
99	7	1.7	104	5	Q9SPM4	Q9SPM4 schistocerc	172	7	1.7	211	2	Q82867	Q82867 rhodovulum
100	7	1.7	104	13	Q9W668	Q9W668 xenopus lae	173	7	1.7	212	16	Q8RP05	Q8RP05 anabarena sp
101	7	1.7	106	9	Q9T111	Q9T111 bacteriophia	174	7	1.7	212	17	Q9UXY4	Q9UXY4 pyrococcus
102	7	1.7	107	5	Q966Y1	Q966Y1 schistocerc	175	7	1.7	213	10	Q9Z081	Q9Z081 arabidopsis
103	7	1.7	113	2	Q9X9S3	Q9X9S3 streptomyc	176	7	1.7	214	16	Q8UDC6	Q8UDC6 agrobacteri
104	7	1.7	124	16	Q910T5	Q910T5 streptomyc	177	7	1.7	215	2	Q934G2	Q934G2 mycobacteri
105	7	1.7	125	10	Q910T3	Q910T3 arabidopsis	178	7	1.7	217	2	Q8KM09	Q8KM09 ruegeria sp
106	7	1.7	131	16	Q8YH06	Q8YH06 bruceella me	179	7	1.7	217	10	Q39826	Q39826 glycine max
107	7	1.7	131	16	Q8G015	Q8G015 bruceella su	180	7	1.7	221	6	Q9SK08	Q9SK08 sus scrofa
108	7	1.7	132	4	Q96B52	Q96B52 homo sapien	181	7	1.7	223	2	Q87773	Q87773 lactobacilli
109	7	1.7	132	4	Q13195	Q13195 homo sapien	182	7	1.7	223	16	Q92547	Q92547 streptomyc
110	7	1.7	132	10	Q9AY50	Q9AY50 chlamydomon	183	7	1.7	225	10	Q949B2	Q949B2 oryza sativ
111	7	1.7	134	2	Q48533	Q48533 leprothrix	184	7	1.7	225	16	Q8FWM8	Q8FWM8 corynebacte
112	7	1.7	135	16	Q9A6E3	Q9A6E3 caulobacter	185	7	1.7	228	8	Q9B661	Q9B661 baiganea ru
113	7	1.7	138	2	Q8KLF1	Q8KLF1 rhizobium e	186	7	1.7	229	2	Q938D9	Q938D9 mycobacteri
114	7	1.7	141	10	Q9AX11	Q9AX11 oryza sativ	187	7	1.7	229	4	Q81Y89	Q81Y89 homo sapien
115	7	1.7	142	2	Q46987	Q46987 escherichia	188	7	1.7	230	16	Q9A513	Q9A513 caulobacter
116	7	1.7	145	16	Q8UB04	Q8UB04 agrobacteri	189	7	1.7	230	16	Q9K1L3	Q9K1L3 streptomyc
117	7	1.7	147	16	O07754	O07754 mycobacteri	190	7	1.7	231	16	B94516	B94516 bacillus su
118	7	1.7	149	10	Q94102	Q94102 oryza sativ	191	7	1.7	233	5	Q9SW10	Q9SW10 crassostrea
119	7	1.7	150	5	Q8T918	Q8T918 drosophila	192	7	1.7	233	10	Q9M224	Q9M224 arabidopsis
120	7	1.7	152	2	Q9XDE5	Q9XDE5 bruceella me	193	7	1.7	234	4	Q9BRM1	Q9BRM1 homo sapien
121	7	1.7	152	5	Q9WEC5	Q9WEC5 drosophila	194	7	1.7	235	13	Q9PTD4	Q9PTD4 oreochromis
122	7	1.7	152	16	Q9ABU8	Q9ABU8 caulobacter	195	7	1.7	235	13	Q9PTD4	Q9PTD4 ictalurys p
123	7	1.7	154	16	Q8ZAX2	Q8ZAX2 yersinia pe	196	7	1.7	236	4	Q9BZ35	Q9BZ35 homo sapien
124	7	1.7	155	10	Q9FT11	Q9FT11 capsicum an	197	7	1.7	240	16	Q8YBR4	Q8YBR4 bruceella me
125	7	1.7	155	10	Q949A8	Q949A8 oryza sativ	198	7	1.7	241	11	Q91ZP6	Q91ZP6 mus musculu
126	7	1.7	155	16	Q8B3G2	Q8B3G2 xanthomonas	199	7	1.7	241	11	Q8BVC5	Q8BVC5 mus musculu
127	7	1.7	156	10	Q94ED8	Q94ED8 oryza sativ	200	7	1.7	242	16	Q9PCD9	Q9PCD9 xyella fas
128	7	1.7	157	17	Q56919	Q56919 pyrococcus	201	7	1.7	243	10	Q82640	Q82640 arabidopsis
129	7	1.7	158	16	Q8UG86	Q8UG86 agrobacteri	202	7	1.7	243	10	Q945W2	Q945W2 oryza sativ
130	7	1.7	163	2	Q8GDE8	Q8GDE8 wolbachia e	203	7	1.7	243	10	Q945X1	Q945X1 oryza sativ
131	7	1.7	163	5	Q10054	Q10054 caenorhabdi	204	7	1.7	244	16	Q8ZFT5	Q8ZFT5 yersinia pe
132	7	1.7	165	2	P77910	P77910 nelsneria g	205	7	1.7	245	16	Q8EANA	Q8EANA shewanella
133	7	1.7	167	16	Q8R7F5	Q8R7F5 thermoaer	206	7	1.7	249	16	Q9POM9	Q9POM9 ureaplaasma
134	7	1.7	168	10	Q9S948	Q9S948 lycopersico	207	7	1.7	253	10	Q9XFM5	Q9XFM5 arabidopsis
135	7	1.7	168	16	Q8ZGZ0	Q8ZGZ0 yersinia pe	208	7	1.7	254	11	Q8C3C7	Q8C3C7 mus musculu
136	7	1.7	169	10	Q9S950	Q9S950 lycopersico	209	7	1.7	254	17	Q8U2D4	Q8U2D4 pyrococcus
137	7	1.7	169	10	Q9S947	Q9S947 lycopersico	210	7	1.7	258	10	Q9AUU0	Q9AUU0 oryza sativ
138	7	1.7	169	10	Q9S8B5	Q9S8B5 fitagaria an	211	7	1.7	258	16	Q92MX0	Q92MX0 rhizobium m
139	7	1.7	169	16	Q8D1P3	Q8D1P3 yersinia pe	212	7	1.7	259	2	Q54168	Q54168 streptomyc
140	7	1.7	171	16	Q8D8P1	Q8D8P1 vibrio vuln	213	7	1.7	260	5	Q95T44	Q95T44 drosophila
141	7	1.7	172	10	Q41119	Q41119 phaseolus v	214	7	1.7	260	16	Q8F545	Q8F545 corynebacte
142	7	1.7	172	16	Q8DHW7	Q8DHW7 synecococc	215	7	1.7	264	5	Q95Q16	Q95Q16 caenorhabdi
143	7	1.7	173	2	Q9F4J9	Q9F4J9 streptococc	216	7	1.7	264	16	Q98K02	Q98K02 rhizobium 1
144	7	1.7	173	4	Q13399	Q13399 homo sapien	217	7	1.7	265	16	Q91L61	Q91L61 streptomyc
145	7	1.7	173	4	Q8N726	Q8N726 homo sapien	218	7	1.7	265	16	Q8PDP4	Q8PDP4 xanthomonas
146	7	1.7	174	17	Q8T021	Q8T021 methanosarc	219	7	1.7	266	2	Q33966	Q33966 saccharopol
147	7	1.7	179	4	Q16360	Q16360 homo sapien	220	7	1.7	267	16	Q9RWS4	Q9RWS4 deinococcus
148	7	1.7	179	16	Q8FLZ2	Q8FLZ2 corynebacte	221	7	1.7	268	12	Q91T10	Q91T10 lupula herp
149	7	1.7	180	16	Q9RZ30	Q9RZ30 deinococcus	222	7	1.7	269	5	Q9NCF3	Q9NCF3 periplaneta
150	7	1.7	181	17	Q26615	Q26615 methanobact	223	7	1.7	269	5	Q9NCF6	Q9NCF6 matricornes
151	7	1.7	182	2	Q93ET2	Q93ET2 cowdria rum	224	7	1.7	270	4	Q96CE9	Q96CE9 homo sapien
152	7	1.7	185	16	Q98717	Q98717 rhizobium 1	225	7	1.7	270	5	Q9NCF1	Q9NCF1 cryptocercu
153	7	1.7	185	16	Q8FLX9	Q8FLX9 corynebacte	226	7	1.7	270	5	Q9NCF5	Q9NCF5 matricornes
154	7	1.7	185	16	Q8RTU3	Q8RTU3 methanosarc	227	7	1.7	270	5	Q9NCF0	Q9NCF0 cryptocercu
155	7	1.7	187	8	Q8HCP3	Q8HCP3 oryza sativ	228	7	1.7	270	5	Q9NCF8	Q9NCF8 neotermes k
156	7	1.7	191	5	Q9N936	Q9N936 leishmania	229	7	1.7	270	5	Q9NCF5	Q9NCF5 polyphaga a
157	7	1.7	191	6	Q28583	Q28583 ovis aries	230	7	1.7	270	5	Q9NCF9	Q9NCF9 cryptocercu
158	7	1.7	193	5	Q20186	Q20186 caenorhabdi	231	7	1.7	270	5	Q9NCF6	Q9NCF6 polyphaga a
159	7	1.7	197	6	Q28584	Q28584 ovis aries	232	7	1.7	270	5	Q9NCF4	Q9NCF4 blattella g
160	7	1.7	198	16	Q9KY89	Q9KY89 streptomyc	233	7	1.7	270	5	Q9NCF7	Q9NCF7 hodotermops
161	7	1.7	202	16	Q9AA57	Q9AA57 caulobacter	234	7	1.7	271	11	Q9RLX2	Q9RLX2 mus musculu
162	7	1.7	203	4	Q8N5R1	Q8N5R1 homo sapien	235	7	1.7	271	11	Q9WV22	Q9WV22 rattus norv

236	7	1.7	272	5	Q81PA7	Q81pa7 drosophila	309	7	1.7	353	2	P70901	P70901 borrelia he
237	7	1.7	273	12	Q84175	Q84175 cotesia con	310	7	1.7	354	10	Q9SYZ8	Q9SYZ8 arabidopsis
238	7	1.7	273	12	Q9YID2	Q9Yid2 cotesia con	311	7	1.7	355	2	Q9ANE3	Q9ANE3 bradyrhizob
239	7	1.7	273	16	Q8UE53	Q8ue53 agrobacteri	312	7	1.7	355	10	Q42853	Q42853 hordeum vul
240	7	1.7	274	5	Q81PA6	Q81pa6 drosophila	313	7	1.7	355	17	Q97W47	Q97W47 sulfobolus
241	7	1.7	275	5	Q8ETX3	Q8etx3 encephalito	314	7	1.7	357	10	Q8H385	Q8H385 oryza sativ
242	7	1.7	275	16	Q9X793	Q9X793 mycobacteri	315	7	1.7	358	2	P70899	P70899 borrelia he
243	7	1.7	275	16	Q98L73	Q98l73 rhizobium 1	316	7	1.7	358	16	Q9RCF4	Q9RCF4 vibrio chol
244	7	1.7	277	16	Q8XRS0	Q8xrs0 ralstonia s	317	7	1.7	360	2	P70905	P70905 borrelia he
245	7	1.7	277	17	Q9HP12	Q9hp12 halobacteri	318	7	1.7	361	10	Q9LZ09	Q9LZ09 arabidopsis
246	7	1.7	280	5	Q81PA5	Q81pa5 drosophila	319	7	1.7	363	12	Q9QJ57	Q9QJ57 human herpe
247	7	1.7	280	11	Q9CRV1	Q9crv1 mus musculu	320	7	1.7	364	10	Q42854	Q42854 hordeum vul
248	7	1.7	281	10	Q8LXK0	Q8lxx0 zea mays (m	321	7	1.7	366	5	Q9NB12	Q9NB12 drosophila
249	7	1.7	281	10	Q8GZ05	Q8gz05 zea mays (m	322	7	1.7	366	5	Q9NB16	Q9NB16 drosophila
250	7	1.7	283	16	Q9PCE4	Q9pce4 xyella fas	323	7	1.7	366	5	Q9NB14	Q9NB14 drosophila
251	7	1.7	284	10	Q9AU05	Q9au05 chlamydomon	324	7	1.7	367	2	Q8RTE5	Q8RTE5 campylobact
252	7	1.7	286	16	Q9HWA9	Q9haw9 pseudomonas	325	7	1.7	367	5	Q81082	Q81082 campylobact
253	7	1.7	287	10	Q949C0	Q949c0 oryza sativ	326	7	1.7	367	16	Q9PID5	Q9PID5 campylobact
254	7	1.7	288	10	Q9SN83	Q9sn83 arabidopsis	327	7	1.7	367	16	Q69922	Q69922 streptomyce
255	7	1.7	289	10	Q8SBA6	Q8sb6 oryza sativ	328	7	1.7	369	16	Q9AJK7	Q9AJK7 caulobacter
256	7	1.7	291	17	Q8ZXJ6	Q8zxj6 pyrobaculum	329	7	1.7	370	5	Q8IIZ9	Q8IIZ9 plasmodium
257	7	1.7	292	16	Q8CT80	Q8ct80 staphylococ	330	7	1.7	370	10	Q9IGT5	Q9IGT5 oryza sativ
258	7	1.7	293	16	P73482	P73482 synchocyst	331	7	1.7	371	3	Q01787	Q01787 planerociae
259	7	1.7	294	16	Q8PSP8	Q8p5p8 xanthomonas	332	7	1.7	372	8	Q953V6	Q953V6 lampyropelti
260	7	1.7	295	16	Q92BG7	Q92bg7 listeria in	333	7	1.7	372	8	Q953R5	Q953R5 lampyropelti
261	7	1.7	296	16	Q8Y6Y4	Q8y6y4 listeria mo	334	7	1.7	372	8	Q953T0	Q953T0 lampyropelti
262	7	1.7	296	16	Q98K10	Q98k10 rhizobium 1	335	7	1.7	374	10	Q9SYZ7	Q9SYZ7 arabidopsis
263	7	1.7	296	16	Q8PH26	Q8ph26 xanthomonas	336	7	1.7	375	16	Q8PR16	Q8PR16 xanthomonas
264	7	1.7	298	16	Q8PMW5	Q8pmw5 xanthomonas	337	7	1.7	376	16	Q92WN8	Q92WN8 rhizobium m
265	7	1.7	301	2	Q9Z6F0	Q9z6f0 mycobacteri	338	7	1.7	377	5	Q81D08	Q81D08 plasmodium
266	7	1.7	302	10	Q8L905	Q8l9j5 arabidopsis	339	7	1.7	378	10	Q8W2W3	Q8W2W3 oryza sativ
267	7	1.7	305	5	Q81PA8	Q81pa8 drosophila	340	7	1.7	378	12	Q65822	Q65822 bovine herp
268	7	1.7	307	10	Q9FUJ6	Q9fj6 arabidopsis	341	7	1.7	379	5	Q81AP3	Q81AP3 plasmodium
269	7	1.7	307	16	Q984P8	Q984p8 rhizobium 1	342	7	1.7	379	16	Q8CY39	Q8CY39 brucella su
270	7	1.7	308	16	Q9KNCS	Q9knc5 vibrio chol	343	7	1.7	380	16	Q98180	Q98180 rhizobium 1
271	7	1.7	309	16	Q8NS10	Q8ns10 coriynabacte	344	7	1.7	381	5	Q813D9	Q813D9 plasmodium
272	7	1.7	309	10	Q8GYS8	Q8gy8 arabidopsis	345	7	1.7	381	11	Q8BNF2	Q8bnf2 mus musculu
273	7	1.7	310	2	Q9FDD6	Q9fda6 selemomonas	346	7	1.7	382	16	Q8E8H1	Q8E8H1 shewanella
274	7	1.7	310	10	Q8S9P6	Q8s9p6 oryza sativ	347	7	1.7	385	5	Q25331	Q25331 leishmania
275	7	1.7	312	16	Q95286	Q95286 homo sapien	348	7	1.7	385	16	Q00745	Q00745 mycobacteri
276	7	1.7	312	16	Q8EP97	Q8ep97 shewanella	349	7	1.7	385	16	Q8XT04	Q8XT04 ralstonia s
277	7	1.7	315	16	Q9K0A9	Q9k0a9 neisseria m	350	7	1.7	386	10	Q8STB8	Q8STB8 oryza sativ
278	7	1.7	315	16	Q9JVB5	Q9jvb5 neisseria m	351	7	1.7	386	10	Q9FUM9	Q9fuh9 zea mays (m
279	7	1.7	315	16	Q9RUS7	Q9rls7 streptomyce	352	7	1.7	388	16	Q8XVU1	Q8xv1 ralstonia s
280	7	1.7	316	16	Q92TA6	Q92ta6 rhizobium m	353	7	1.7	389	2	Q9ZNC9	Q9znc9 pseudomonas
281	7	1.7	317	10	Q93WY7	Q93wy7 nicotiana t	354	7	1.7	390	2	Q9APV7	Q9ap7 treponema d
282	7	1.7	317	10	Q93WY8	Q93wy8 nicotiana t	355	7	1.7	392	2	Q9F3Y3	Q9f3y3 pasteurella
283	7	1.7	317	16	Q8ZAT8	Q8zat8 yerisina pe	356	7	1.7	392	16	Q8PEB7	Q8peb7 xanthomonas
284	7	1.7	319	10	Q38817	Q38817 arabidopsis	357	7	1.7	393	16	Q50993	Q50993 neisseria g
285	7	1.7	320	16	Q8NPV3	Q8npv3 coriynabacte	358	7	1.7	393	16	Q9JZP6	Q9jzp6 neisseria m
286	7	1.7	320	17	Q8TL26	Q8tl26 methanosarc	359	7	1.7	393	16	Q50532	Q50532 streptomyce
287	7	1.7	324	5	Q9XTY9	Q9xty9 caenorhabdi	360	7	1.7	394	5	Q9NJS6	Q9njs6 aplysia kur
288	7	1.7	324	16	Q9PHI3	Q9ph3 xyella fas	361	7	1.7	394	5	Q9NHP3	Q9nhp3 aplysia cal
289	7	1.7	325	10	Q43751	Q43751 capsicum an	362	7	1.7	395	16	Q9ZBN5	Q9zbn5 streptomyce
290	7	1.7	325	11	Q8BRB6	Q8brb6 mus musculu	363	7	1.7	396	2	Q9ZH25	Q9zh25 pseudomonas
291	7	1.7	330	2	Q8GR66	Q8gre6 borrelia du	364	7	1.7	396	16	Q9RG60	Q9rg60 pseudomonas
292	7	1.7	330	16	Q9PUH4	Q9puh4 chlamydia m	365	7	1.7	398	2	Q9EXN8	Q9exn8 escherichia
293	7	1.7	331	16	Q9ACU4	Q9acu4 streptomyce	366	7	1.7	398	2	Q9ZT83	Q9zt83 proteus mir
294	7	1.7	332	4	Q8N2U1	Q8n2u1 homo sapien	367	7	1.7	398	16	Q8Y3G9	Q8y3g9 ralstonia s
295	7	1.7	332	11	Q8COK5	Q8cok5 mus musculu	368	7	1.7	398	16	Q8PHS0	Q8phs0 xanthomonas
296	7	1.7	333	16	Q8FUD0	Q8fud0 coriynabacte	369	7	1.7	398	17	Q30103	Q30103 archaeoglob
297	7	1.7	334	16	Q8YU05	Q8yu05 ralstonia s	370	7	1.7	399	2	Q87935	Q87935 burkholderi
298	7	1.7	336	10	Q8S7Y4	Q8s7y4 oryza sativ	371	7	1.7	399	16	Q8RR04	Q8rr04 coriynabacte
299	7	1.7	336	16	Q9ZVM9	Q9zvm9 rhizobium m	372	7	1.7	400	2	Q9ZAI1	Q9zai1 pasteurella
300	7	1.7	339	17	Q8XXA2	Q8xxa2 ralstonia s	373	7	1.7	400	2	Q9R758	Q9r758 pasteurella
301	7	1.7	340	16	Q8TIM4	Q8tim4 methanosarc	374	7	1.7	400	16	Q8G7Y2	Q8g7y2 bifidobacte
302	7	1.7	341	16	Q98GK4	Q98gk4 rhizobium 1	375	7	1.7	401	16	Q8CUM3	Q8cwm3 streptomyce
303	7	1.7	345	2	Q8GRD4	Q8grd4 borrelia du	376	7	1.7	402	16	Q98813	Q98813 rhizobium 1
304	7	1.7	346	16	Q9AKS24	Q9aks24 ralstonia s	377	7	1.7	402	16	Q69865	Q69865 streptomyce
305	7	1.7	348	16	Q9AK41	Q9ak41 streptomyce	378	7	1.7	402	16	Q8CNA1	Q8cna1 staphylococ
306	7	1.7	348	17	Q8TZR7	Q8tzt7 pyrococcus	379	7	1.7	403	16	Q9JUT2	Q9jut2 neisseria m
307	7	1.7	349	2	Q9F811	Q9f811 ewinia amy	380	7	1.7	404	16	Q06263	Q06263 mycobacteri
308	7	1.7	350	16	Q9X8V8	Q9x8v8 streptomyce	381	7	1.7	404	16	Q8XZ64	Q8xz64 ralstonia s

382	1.7	405	2	Q934L1	Q93411 vibrio salm	455	7	1.7	487	4	Q95172	Q95172 homo sapien
383	1.7	405	2	Q52225	Q52225 agrobacteri	456	7	1.7	488	10	Q22298	Q22298 citrus sine
384	1.7	406	16	Q8DB09	Q8db09 vibrio vuln	457	7	1.7	488	11	Q8K3V8	Q8K3V8 rattus norv
385	1.7	407	2	Q92619	Q92619 anaplasm m	458	7	1.7	489	10	Q42872	Q42872 lycopersico
386	1.7	407	10	Q9S7J3	Q9S7J3 weilateschia	459	7	1.7	489	10	Q49296	Q49296 arabidopsis
387	1.7	409	2	Q44092	Q44092 anaplasm m	460	7	1.7	489	10	Q93W21	Q93W21 nicotiana t
388	1.7	409	10	Q8H653	Q8H653 oryza sativ	461	7	1.7	490	16	Q8XW69	Q8XW69 ralsconia s
389	1.7	409	16	Q69974	Q69974 streptomyce	462	7	1.7	490	10	Q9CAC1	Q9CAC1 arabidopsis
390	1.7	410	2	Q8RSK4	Q8RSK4 uncultured	463	7	1.7	492	10	Q23697	Q23697 arabidopsis
391	1.7	410	2	Q9R965	Q9R965 anaplasm m	464	7	1.7	493	10	Q23696	Q23696 arabidopsis
392	1.7	410	16	Q92JW7	Q92JW7 rhizobium m	465	7	1.7	493	10	Q9AYW9	Q9AYW9 atriplex le
393	1.7	411	2	Q9RH56	Q9RH56 bradyrhizob	466	7	1.7	493	10	Q9SVJ3	Q9SVJ3 arabidopsis
394	1.7	411	17	Q8PYR2	Q8PYR2 methanosarc	467	7	1.7	494	10	Q9SVJ4	Q9SVJ4 arabidopsis
395	1.7	417	16	Q91478	Q91478 pseudomonas	468	7	1.7	494	10	Q9X1Y8	Q9X1Y8 populus alb
396	1.7	417	16	Q9RY67	Q9RY67 deinococcus	469	7	1.7	494	10	Q40763	Q40763 populus alb
397	1.7	418	17	Q9YFO0	Q9YFO0 aeropyrum p	470	7	1.7	494	10	Q9AVT5	Q9AVT5 populus alb
398	1.7	420	17	Q9HNG8	Q9HNG8 halobacteri	471	7	1.7	496	10	Q81403	Q81403 fragaria an
399	1.7	422	16	Q8PRB8	Q8PRB8 xanthomonas	472	7	1.7	496	10	Q92T50	Q92T50 fragaria an
400	1.7	423	2	Q9Z620	Q9Z620 anaplasm m	473	7	1.7	496	10	Q9ZTL0	Q9ZTL0 fragaria an
401	1.7	431	17	Q9H161	Q9H161 thermoplasm	474	7	1.7	496	10	Q9S8T6	Q9S8T6 fragaria an
402	1.7	433	5	Q96819	Q96819 trypanosoma	475	7	1.7	497	5	Q9GRR2	Q9GRR2 lycopersico
403	1.7	433	12	Q9WT62	Q9WT62 human herpe	476	7	1.7	497	10	Q82473	Q82473 lycopersico
404	1.7	435	5	Q817C8	Q817C8 ciona intes	477	7	1.7	497	10	Q96546	Q96546 capsicum an
405	1.7	436	2	Q61556	Q61556 streptomyce	478	7	1.7	497	10	Q9SML6	Q9SML6 capsicum an
406	1.7	436	16	Q8XAF6	Q8XAF6 escherichia	479	7	1.7	497	10	P94114	P94114 prunus pers
407	1.7	436	16	Q8FDC6	Q8FDC6 escherichia	480	7	1.7	498	4	Q9NPE8	Q9NPE8 homo sapien
408	1.7	437	2	Q9F1V8	Q9F1V8 streptomyce	481	7	1.7	499	10	Q8LO92	Q8LO92 oryza sativ
409	1.7	440	12	Q65588	Q65588 bovine hairp	482	7	1.7	501	2	Q8KY30	Q8KY30 streptomyce
410	1.7	441	10	Q94C38	Q94C38 arabidopsis	483	7	1.7	501	10	Q64949	Q64949 arabidopsis
411	1.7	441	16	Q9RZV2	Q9RZV2 deinococcus	484	7	1.7	501	10	Q9SRK3	Q9SRK3 arabidopsis
412	1.7	443	11	Q91ZD7	Q91ZD7 rattus norv	485	7	1.7	502	16	Q8ZGP5	Q8ZGP5 yerania pe
413	1.7	444	5	Q9V839	Q9V839 drosophila	486	7	1.7	504	10	Q9LDF0	Q9LDF0 oryza sativ
414	1.7	444	16	Q8XZ44	Q8XZ44 ralsconia s	487	7	1.7	505	10	Q22297	Q22297 citrus sine
415	1.7	447	5	Q9NCE2	Q9NCE2 panesthia c	488	7	1.7	506	16	Q9SSU7	Q9SSU7 distus sine
416	1.7	447	5	Q81FUE	Q81FUE mastoetermes	489	7	1.7	506	16	Q8PSW8	Q8PSW8 xanthomonas
417	1.7	447	5	Q81FUA	Q81FUA mastoetermes	490	7	1.7	507	10	Q9ATV1	Q9ATV1 lolium rigi
418	1.7	447	16	Q9KMX3	Q9KMX3 vibrio chol	491	7	1.7	507	16	Q8CKR6	Q8CKR6 streptomyce
419	1.7	447	16	Q9K935	Q9K935 bacillus ha	492	7	1.7	508	10	Q43149	Q43149 sambucus ni
420	1.7	448	2	Q93HP4	Q93HP4 streptomyce	493	7	1.7	508	16	Q86591	Q86591 streptomyce
421	1.7	448	5	Q9BMC7	Q9BMC7 coproetermes	494	7	1.7	509	16	Q8Y3F0	Q8Y3F0 ralsconia s
422	1.7	448	5	Q77045	Q77045 nasutiterme	495	7	1.7	510	5	Q44018	Q44018 leishmania
423	1.7	448	5	Q77044	Q77044 nasutiterme	496	7	1.7	510	10	Q9ATV2	Q9ATV2 lolium rigi
424	1.7	448	5	Q81FUE3	Q81FUE3 mastoetermes	497	7	1.7	510	10	Q42875	Q42875 lycopersico
425	1.7	448	5	Q81FUE2	Q81FUE2 mastoetermes	498	7	1.7	510	10	Q64401	Q64401 pinus radia
426	1.7	449	5	Q81FUE5	Q81FUE5 mastoetermes	499	7	1.7	511	10	Q9LE72	Q9LE72 oryza sativ
427	1.7	449	16	Q8RCG1	Q8RCG1 thermococ	500	7	1.7	514	4	Q9BWJ3	Q9BWJ3 homo sapien
428	1.7	450	5	Q9NCE3	Q9NCE3 panesthia c	501	7	1.7	515	10	Q64402	Q64402 pinus radia
429	1.7	456	2	Q9AF07	Q9AF07 frankia sp.	502	7	1.7	516	10	Q81416	Q81416 arabidopsis
430	1.7	456	5	Q9N9U6	Q9N9U6 leishmania	503	7	1.7	519	5	Q8MRX3	Q8MRX3 drosophila
431	1.7	460	10	Q8L1Y8	Q8L1Y8 oryza sativ	504	7	1.7	525	16	Q8RG15	Q8RG15 fusobacteri
432	1.7	461	5	Q8T7Y2	Q8T7Y2 bombyx mori	505	7	1.7	529	2	Q50747	Q50747 mycobacteri
433	1.7	464	10	Q9AHY3	Q9AHY3 oryza sativ	506	7	1.7	530	3	Q14430	Q14430 cryphonectr
434	1.7	465	4	Q96GY4	Q96GY4 homo sapien	507	7	1.7	530	16	Q05813	Q05813 mycobacteri
435	1.7	469	5	Q9Y0W2	Q9Y0W2 cherax quad	508	7	1.7	531	16	Q98D82	Q98D82 rhizobium l
436	1.7	470	5	Q9V3E1	Q9V3E1 drosophila	509	7	1.7	532	2	Q9F7M5	Q9F7M5 uncultured
437	1.7	470	5	Q8SVF2	Q8SVF2 drosophila	510	7	1.7	533	16	Q53903	Q53903 streptomyce
438	1.7	471	10	Q8LAG1	Q8LAG1 arabidopsis	511	7	1.7	535	16	Q9CBT1	Q9CBT1 mycobacteri
439	1.7	471	10	Q8VZES	Q8VZES arabidopsis	512	7	1.7	536	2	Q8RLC1	Q8RLC1 comamonas a
440	1.7	473	2	Q9RQ6E	Q9RQ6E cellulomona	513	7	1.7	538	5	Q9Y115	Q9Y115 drosophila
441	1.7	477	4	Q8NAT2	Q8NAT2 homo sapien	514	7	1.7	539	16	Q9RK08	Q9RK08 streptomyce
442	1.7	479	10	Q04972	Q04972 lycopersico	515	7	1.7	540	17	Q8TPU7	Q8TPU7 methanosarc
443	1.7	479	10	Q9SUS0	Q9SUS0 arabidopsis	516	7	1.7	541	10	Q9SGR6	Q9SGR6 arabidopsis
444	1.7	480	10	Q9S290	Q9S290 arabidopsis	517	7	1.7	541	10	Q93ZU4	Q93ZU4 arabidopsis
445	1.7	481	10	Q9ZVZ7	Q9ZVZ7 arabidopsis	518	7	1.7	541	10	Q8L790	Q8L790 arabidopsis
446	1.7	482	16	Q8CJX6	Q8CJX6 streptomyce	519	7	1.7	543	11	Q8BNH8	Q8BNH8 mus musculu
447	1.7	484	2	Q91AV5	Q91AV5 thermomonos	520	7	1.7	543	16	Q69831	Q69831 streptomyce
448	1.7	484	10	Q9C9H5	Q9C9H5 arabidopsis	521	7	1.7	546	16	Q986M2	Q986M2 rhizobium l
449	1.7	484	10	Q8LDB8	Q8LDB8 arabidopsis	522	7	1.7	551	16	Q9HUD4	Q9HUD4 pseudomonas
450	1.7	485	10	Q96547	Q96547 capsicum an	523	7	1.7	552	3	Q96TJ9	Q96TJ9 utromyces fa
451	1.7	486	10	Q41012	Q41012 pisum sativ	524	7	1.7	555	11	Q8C3W8	Q8C3W8 mus musculu
452	1.7	486	10	Q9FK01	Q9FK01 arabidopsis	525	7	1.7	555	11	Q8C3C6	Q8C3C6 mus musculu
453	1.7	487	4	Q9UGA3	Q9UGA3 homo sapien	526	7	1.7	555	11	Q8BH92	Q8BH92 mus musculu
454	1.7	487	4	Q9UNW1	Q9UNW1 homo sapien	527	7	1.7	559	11	Q8R136	Q8R136 mus musculu

528	7	1.7	559	17	Q9HRT5	Q9HRT5 halobacteri	601	7	1.7	773	16	Q9L1R8	Q9L1R8 streptomyc
529	7	1.7	571	10	O04250	O04250 arabidopsis	602	7	1.7	777	5	O61366	O61366 drosophila
530	7	1.7	572	3	O42728	O42728 rhodospirid	603	7	1.7	782	16	Q988Y2	Q988Y2 rhizobium 1
531	7	1.7	574	16	Q91246	Q91246 pseudomonas	604	7	1.7	786	5	O8T107	O8T107 bombax mori
532	7	1.7	581	2	O51863	O51863 escherichia	605	7	1.7	786	5	O24014	O24014 dictyosteli
533	7	1.7	581	2	Q9S6H9	Q9S6H9 escherichia	606	7	1.7	789	16	Q8UGX6	Q8UGX6 agrobacteri
534	7	1.7	582	5	O62146	O62146 caenorhabdi	607	7	1.7	802	4	Q9BQH6	Q9BQH6 homo sapien
535	7	1.7	582	16	O67141	O67141 aquilex aeo	608	7	1.7	813	2	O8GFJ3	Q8GFJ3 streptomyc
536	7	1.7	584	17	Q9HON3	Q9HON3 halobacteri	609	7	1.7	827	10	O8S5J7	O8S5J7 oryza sativ
537	7	1.7	586	4	Q8NA76	Q8NA76 homo sapien	610	7	1.7	840	2	O44003	O44003 ralstonia m
538	7	1.7	586	10	Q81WD5	Q81WD5 homo sapien	611	7	1.7	850	16	Q8XZD5	Q8XZD5 ralstonia s
539	7	1.7	586	10	Q949C3	Q949C3 oryza sativ	612	7	1.7	859	15	Q8Q850	Q8Q850 human immun
540	7	1.7	587	16	Q8DCU4	Q8DCU4 vibrio vuln	613	7	1.7	860	2	Q9L428	Q9L428 bordetella
541	7	1.7	592	16	Q8PGR6	Q8PGR6 xanthomonas	614	7	1.7	870	5	O8MSY6	O8MSY6 drosophila
542	7	1.7	594	3	Q9P6E2	Q9P6E2 neurospora	615	7	1.7	875	16	Q8RBA3	Q8RBA3 thermocaneer
543	7	1.7	594	16	Q8PER6	Q8PER6 xanthomonas	616	7	1.7	884	10	Q94FE0	Q94FE0 dichantheli
544	7	1.7	595	4	Q9ULP9	Q9ULP9 homo sapien	617	7	1.7	884	10	Q8GV57	Q8GV57 oryza sativ
545	7	1.7	597	5	O02137	O02137 caenorhabdi	618	7	1.7	887	1	O50743	Q50743 methanobact
546	7	1.7	600	16	Q9H2M1	Q9H2M1 pseudomonas	619	7	1.7	891	11	Q8VHG2	Q8VHG2 mus musculu
547	7	1.7	600	16	Q8G5G9	Q8G5G9 bifidobacte	620	7	1.7	893	2	O8GCW9	O8GCW9 chromobacte
548	7	1.7	605	11	Q8BH65	Q8BH65 mus musculu	621	7	1.7	895	2	O52507	O52507 serratia 1i
549	7	1.7	610	2	Q9S5X1	Q9S5X1 streptomyc	622	7	1.7	900	5	Q9V7S2	Q9V7S2 drosophila
550	7	1.7	610	2	O69311	O69311 streptomyc	623	7	1.7	901	17	Q9HRV9	Q9HRV9 halobacteri
551	7	1.7	610	16	Q86826	Q86826 streptomyc	624	7	1.7	917	3	Q9HGS0	Q9HGS0 botrytis ci
552	7	1.7	612	16	Q8FEP8	Q8FEP8 escherichia	625	7	1.7	918	16	Q8Y1A4	Q8Y1A4 ralstonia s
553	7	1.7	615	4	Q9H7G2	Q9H7G2 homo sapien	626	7	1.7	941	16	O9Z1W5	Q9Z1W5 streptomyc
554	7	1.7	615	4	Q96CM8	Q96CM8 homo sapien	627	7	1.7	947	5	O20625	Q20625 caenorhabdi
555	7	1.7	615	11	Q8VCM8	Q8VCM8 mus musculu	628	7	1.7	955	16	Q8XW19	Q8XW19 ralstonia s
556	7	1.7	617	16	Q92NX3	Q92NX3 rhizobium m	629	7	1.7	959	16	Q8P524	Q8P524 xanthomonas
557	7	1.7	618	11	Q8C5Q0	Q8C5Q0 mus musculu	630	7	1.7	965	5	Q9BHC2	Q9BHC2 leishmania
558	7	1.7	619	2	O32374	O32374 clostridium	631	7	1.7	970	12	Q8Q206	Q8Q206 rana tigrin
559	7	1.7	620	10	Q8RMR6	Q8RMR6 triticum ae	632	7	1.7	970	16	Q8G6Y0	Q8G6Y0 bifidobacte
560	7	1.7	621	10	Q9MAY8	Q9MAY8 hordeum vul	633	7	1.7	978	16	Q8PHB6	Q8PHB6 xanthomonas
561	7	1.7	623	10	Q9FVQ2	Q9FVQ2 arabidopsis	634	7	1.7	978	16	O8PE21	O8PE21 xanthomonas
562	7	1.7	624	16	Q91237	Q91237 pseudomonas	635	7	1.7	980	10	Q8WKX8	Q8WKX8 oryza sativ
563	7	1.7	624	16	Q8P4H2	Q8P4H2 xanthomonas	636	7	1.7	981	16	Q8XS09	Q8XS09 ralstonia s
564	7	1.7	625	16	Q8PG33	Q8PG33 xanthomonas	637	7	1.7	987	10	Q94DR0	Q94DR0 oryza sativ
565	7	1.7	626	5	Q9NDD1	Q9NDD1 leishmania	638	7	1.7	993	11	O8CC22	O8CC22 mus musculu
566	7	1.7	626	16	Q8XW42	Q8XW42 ralstonia s	639	7	1.7	995	4	O81YU1	O81YU1 homo sapien
567	7	1.7	627	10	Q9M995	Q9M995 arabidopsis	640	7	1.7	997	16	Q98GZ1	Q98GZ1 rhizobium 1
568	7	1.7	631	16	O86724	O86724 streptomyc	641	7	1.7	1004	10	Q8H384	Q8H384 oryza sativ
569	7	1.7	635	2	O30348	O30348 ralstonia s	642	7	1.7	1005	16	Q9HTE6	Q9HTE6 pseudomonas
570	7	1.7	640	5	Q24256	Q24256 drosophila	643	7	1.7	1008	11	Q9DBV8	Q9DBV8 mus musculu
571	7	1.7	645	5	Q9VX55	Q9VX55 drosophila	644	7	1.7	1018	16	Q8EDB9	Q8EDB9 shewanella
572	7	1.7	649	3	O43069	O43069 schizosacch	645	7	1.7	1030	5	O45541	O45541 caenorhabdi
573	7	1.7	658	16	Q8BE51	Q8BE51 shewanella	646	7	1.7	1041	4	Q96P70	Q96P70 homo sapien
574	7	1.7	660	10	Q941T0	Q941T0 oryza sativ	647	7	1.7	1041	16	Q8PHD3	Q8PHD3 homo sapien
575	7	1.7	662	5	Q9V9F7	Q9V9F7 drosophila	648	7	1.7	1041	16	Q8PE13	Q8PE13 xanthomonas
576	7	1.7	662	16	Q93RX7	Q93RX7 streptomyc	649	7	1.7	1044	4	Q961W3	Q961W3 homo sapien
577	7	1.7	668	5	Q9BLR2	Q9BLR2 leishmania	650	7	1.7	1046	10	Q8LHJ8	Q8LHJ8 xanthomonas
578	7	1.7	673	3	O74228	O74228 emericella	651	7	1.7	1053	10	Q9PWS3	Q9PWS3 chlamydomon
579	7	1.7	674	4	Q96SUC	Q96SUC homo sapien	652	7	1.7	1070	16	O53165	O53165 synechocyst
580	7	1.7	674	4	Q8NCG9	Q8NCG9 homo sapien	653	7	1.7	1070	16	Q8F6Q0	Q8F6Q0 leptospira
581	7	1.7	677	10	Q9M2F5	Q9M2F5 arabidopsis	654	7	1.7	1097	3	O13592	O13592 saccharomyc
582	7	1.7	677	16	Q8EJN8	Q8EJN8 shewanella	655	7	1.7	1097	16	Q8PP99	Q8PP99 xanthomonas
583	7	1.7	681	16	Q9KX65	Q9KX65 streptomyc	656	7	1.7	1099	2	O68840	O68840 bacillus st
584	7	1.7	682	16	Q8XW42	Q8XW42 ralstonia s	657	7	1.7	1130	4	Q96E12	Q96E12 homo sapien
585	7	1.7	687	2	O60103	O60103 xanthobact	658	7	1.7	1136	2	Q9X686	Q9X686 salmonella
586	7	1.7	689	4	O81ZV0	O81ZV0 homo sapien	659	7	1.7	1137	2	Q9AGH9	Q9AGH9 pseudomonas
587	7	1.7	689	11	Q8CJ60	Q8CJ60 mus musculu	660	7	1.7	1157	4	Q96JUK7	Q96JUK7 homo sapien
588	7	1.7	689	16	Q92N56	Q92N56 rhizobium m	661	7	1.7	1173	4	O81X21	O81X21 homo sapien
589	7	1.7	691	5	Q9VZE7	Q9VZE7 drosophila	662	7	1.7	1207	16	Q9A797	Q9A797 caulobacter
590	7	1.7	701	2	Q9KM03	Q9KM03 pseudomonas	663	7	1.7	1213	4	Q92923	Q92923 homo sapien
591	7	1.7	721	16	O8XB65	O8XB65 escherichia	664	7	1.7	1214	4	O8TAQ2	O8TAQ2 homo sapien
592	7	1.7	721	16	O8FA82	O8FA82 escherichia	665	7	1.7	1215	5	O20626	O20626 caenorhabdi
593	7	1.7	725	2	O8KY01	O8KY01 rhodospendo	666	7	1.7	1237	4	O75147	O75147 homo sapien
594	7	1.7	728	16	Q8PH89	Q8PH89 xanthomonas	667	7	1.7	1257	5	Q9VKZ8	Q9VKZ8 drosophila
595	7	1.7	729	12	Q9E114	Q9E114 melalegri h	668	7	1.7	1258	17	O8ZYM3	O8ZYM3 pyrobaculum
596	7	1.7	729	12	Q9DP88	Q9DP88 melalegri h	669	7	1.7	1345	16	Q9Z512	Q9Z512 mycobacteri
597	7	1.7	735	10	O81954	O81954 chlamydomon	670	7	1.7	1365	2	O8KWD2	O8KWD2 ruergeria sp
598	7	1.7	738	5	O02402	O02402 pinctada fu	671	7	1.7	1371	16	O8XQ42	O8XQ42 ralstonia s
599	7	1.7	745	16	Q8G6V5	Q8G6V5 bifidobacte	672	7	1.7	1388	5	O8WCJ6	O8WCJ6 leishmania
600	7	1.7	762	10	Q9SVV7	Q9SVV7 arabidopsis	673	7	1.7	1432	11	Q8CHD1	Q8CHD1 mus musculu

674	1.7	1530	16	Q9KPU4	Q9KPU4 vibrio chol	747	6	1.5	67	16	Q8EPF9	Q8EPF9 shewanella
675	1.7	1605	2	Q9X6M3	Q9X6M3 salmonella	748	6	1.5	69	2	Q54580	Q54580 streptococc
676	1.7	1734	11	Q6O592	Q6O592 mus musculu	749	6	1.5	69	10	Q94K36	Q94K36 arabadopsis
677	1.7	1787	16	Q8Y422	Q8Y422 listeria mo	750	6	1.5	69	12	Q91LD9	Q91LD9 white spot
678	1.7	1788	16	Q92F16	Q92F16 listeria in	751	6	1.5	70	9	Q8SCP3	Q8SCP3 pseudomonas
679	1.7	1963	16	Q8EGK2	Q8EGK2 shewanella	752	6	1.5	70	10	Q8LE23	Q8LE23 arabadopsis
680	1.7	2030	5	Q9VWN8	Q9VWN8 dirosophila	753	6	1.5	70	16	Q8RE06	Q8RE06 fusobacteri
681	1.7	2055	2	Q85472	Q85472 abiotrophia	754	6	1.5	71	10	Q8GXU7	Q8GXU7 arabadopsis
682	1.7	2139	5	Q9VU09	Q9VU09 dirosophila	755	6	1.5	71	16	Q8JDE1	Q8JDE1 rhizobium 1
683	1.7	2229	16	Q9EMP5	Q9EMP5 streptomyce	756	6	1.5	71	16	Q97HS9	Q97HS9 clostridium
684	1.7	2335	2	Q93HH9	Q93HH9 streptomyce	757	6	1.5	72	4	Q8H678	Q8H678 oryza sativ
685	1.7	2656	5	Q9GNJ3	Q9GNJ3 paracentrot	758	6	1.5	72	10	P72995	P72995 synechocyst
686	1.7	2736	5	Q9NKG6	Q9NKG6 leishmania	759	6	1.5	72	16	Q39543	Q39543 hepatitis g
687	1.7	2946	10	Q64634	Q64634 arabadopsis	760	6	1.5	73	12	Q9PME7	Q9PME7 campylobact
688	1.7	2957	5	Q61845	Q61845 caenorhabdi	761	6	1.5	73	16	Q92M08	Q92M08 rhizobium m
689	1.7	2993	16	Q8NMS0	Q8NMS0 corynebacte	762	6	1.5	73	16	Q92M05	Q92M05 rhizobium m
690	1.7	2998	5	Q8MXK6	Q8MXK6 leishmania	763	6	1.5	73	16	Q8R5T5	Q8R5T5 thermoaer
691	1.7	2998	5	Q9NKS4	Q9NKS4 leishmania	764	6	1.5	73	16	Q8P0S1	Q8P0S1 streptococc
692	1.7	3084	12	Q8U211	Q8U211 pseudorabie	765	6	1.5	74	2	Q50168	Q50168 mycobacteri
693	1.7	3124	4	Q96L91	Q96L91 homo sapien	766	6	1.5	74	2	Q8LID2	Q8LID2 streptomyce
694	1.7	3247	12	Q65553	Q65553 bovine herp	767	6	1.5	75	16	Q8CTP3	Q8CTP3 staphylococ
695	1.7	3346	12	Q66666	Q66666 equine herp	768	6	1.5	75	16	Q98110	Q98110 rhizobium 1
696	1.7	3353	16	Q91791	Q91791 pseudomonas	769	6	1.5	76	16	Q8SBJ5	Q8SBJ5 bacterioph
697	1.7	3591	2	Q9KIE1	Q9KIE1 streptomyce	770	6	1.5	77	9	Q8SBJ5	Q8SBJ5 bacterioph
698	1.7	3624	16	Q8Z411	Q8Z411 salmonella	771	6	1.5	77	16	Q9PLY2	Q9PLY2 campylobact
699	1.7	3982	2	Q9J156	Q9J156 bacillus su	772	6	1.5	77	16	Q8FYU0	Q8FYU0 brucella su
700	1.7	4124	5	Q97218	Q97218 leishmania	773	6	1.5	77	17	Q8PYL1	Q8PYL1 methanosa
701	1.7	5627	16	Q91120	Q91120 pseudomonas	774	6	1.5	78	16	Q9CLJ2	Q9CLJ2 pasteurella
702	1.7	9477	2	Q914X3	Q914X3 streptomyce	775	6	1.5	78	16	Q9BM09	Q9BM09 rhizobium 1
703	1.7	10917	2	Q9JNM6	Q9JNM6 streptomyce	776	6	1.5	79	16	Q9RK06	Q9RK06 streptomyce
704	1.5	31	2	Q9JMV2	Q9JMV2 escherichia	777	6	1.5	80	4	Q8TF08	Q8TF08 homo sapien
705	1.5	34	16	Q8KAV8	Q8KAV8 chlorobium	778	6	1.5	80	5	Q8ISR3	Q8ISR3 spodoptera
706	1.5	35	2	Q8KYN5	Q8KYN5 bacillus an	779	6	1.5	80	16	P71992	P71992 mycobacteri
707	1.5	35	16	Q8VJY6	Q8VJY6 mycobacteri	780	6	1.5	81	5	Q77134	Q77134 dirosophila
708	1.5	35	16	Q9FBM7	Q9FBM7 streptomyce	781	6	1.5	81	16	Q9KT78	Q9KT78 vibrio chol
709	1.5	36	6	Q9JTP8	Q9JTP8 leontopithe	782	6	1.5	81	16	Q8X807	Q8X807 escherichia
710	1.5	37	2	Q9EY28	Q9EY28 photobacter	783	6	1.5	81	16	Q8D897	Q8D897 vibrio vuln
711	1.5	40	16	Q8E168	Q8E168 shewanella	784	6	1.5	82	2	Q99Q71	Q99Q71 vibrio harv
712	1.5	42	2	Q9R730	Q9R730 xanthomonas	785	6	1.5	82	2	Q9AJG2	Q9AJG2 vibrio camp
713	1.5	42	10	Q9SQ83	Q9SQ83 arabadopsis	786	6	1.5	82	2	Q54721	Q54721 spirulina p
714	1.5	43	16	Q8XW38	Q8XW38 raiistonia s	787	6	1.5	82	5	Q96486	Q96486 plasmodium
715	1.5	49	6	Q28871	Q28871 canis fam11	788	6	1.5	82	9	Q38378	Q38378 bacterioph
716	1.5	50	16	Q8P0L0	Q8P0L0 streptococc	789	6	1.5	82	9	Q8SBK0	Q8SBK0 bacterioph
717	1.5	50	16	Q8G0A7	Q8G0A7 brucella su	790	6	1.5	83	2	Q85410	Q85410 pseudomonas
718	1.5	51	1	Q9C4M2	Q9C4M2 sulfolobus	791	6	1.5	83	2	Q07490	Q07490 yersinia en
719	1.5	51	1	Q9S505	Q9S505 homo sapien	792	6	1.5	83	2	Q85417	Q85417 pseudomonas
720	1.5	51	16	Q98G54	Q98G54 rhizobium 1	793	6	1.5	83	10	Q80697	Q80697 arabadopsis
721	1.5	54	4	Q9H2D8	Q9H2D8 homo sapien	794	6	1.5	83	16	Q9A2W0	Q9A2W0 caulobacter
722	1.5	55	6	Q8S086	Q8S086 salmimi sci	795	6	1.5	83	16	Q8Z0Q0	Q8Z0Q0 salmonella
723	1.5	55	6	Q8S0B4	Q8S0B4 ateles fusc	796	6	1.5	83	16	Q8Z886	Q8Z886 salmonella
724	1.5	55	6	Q8S0H2	Q8S0H2 ateles belz	797	6	1.5	83	16	Q69938	Q69938 streptomyce
725	1.5	55	6	Q8S0Q8	Q8S0Q8 leontopithe	798	6	1.5	85	5	Q9BHC3	Q9BHC3 leishmania
726	1.5	55	6	Q8S0S5	Q8S0S5 lemur varie	799	6	1.5	86	2	Q9S4M5	Q9S4M5 escherichia
727	1.5	55	16	Q9PFY6	Q9PFY6 xyfella fas	800	6	1.5	86	9	Q9MCR8	Q9MCR8 bacterioph
728	1.5	56	4	Q16750	Q16750 homo sapien	801	6	1.5	86	16	Q92PD9	Q92PD9 rhizobium m
729	1.5	59	2	Q8VTH2	Q8VTH2 helicobacte	802	6	1.5	87	16	Q8PBL7	Q8PBL7 xanthomonas
730	1.5	59	16	Q8ZRS8	Q8ZRS8 salmonella	803	6	1.5	87	16	Q8NUC8	Q8NUC8 corynebacte
731	1.5	59	16	Q9ZK15	Q9ZK15 helicobacte	804	6	1.5	87	16	Q8RUK5	Q8RUK5 corynebacte
732	1.5	60	16	Q8N2T7	Q8N2T7 streptococc	805	6	1.5	88	2	Q56971	Q56971 yersinia pe
733	1.5	60	16	Q8K6X1	Q8K6X1 streptococc	806	6	1.5	88	16	Q8FPU4	Q8FPU4 escherichia
734	1.5	61	12	Q65583	Q65583 bovine herp	807	6	1.5	89	2	Q33552	Q33552 rhodobacter
735	1.5	61	14	Q93156	Q93156 uncultured	808	6	1.5	89	7	Q95956	Q95956 homo sapien
736	1.5	61	16	Q8XNV2	Q8XNV2 clostridium	809	6	1.5	89	7	Q77959	Q77959 homo sapien
737	1.5	61	16	Q8Y106	Q8Y106 brucella me	810	6	1.5	90	9	Q9MCC7	Q9MCC7 bacterioph
738	1.5	62	16	Q8YLD8	Q8YLD8 anabaena sp	811	6	1.5	91	2	Q48265	Q48265 haemophilus
739	1.5	62	16	Q8K799	Q8K799 streptococc	812	6	1.5	91	2	Q85240	Q85240 yersinia en
740	1.5	62	16	Q8EAG4	Q8EAG4 shewanella	813	6	1.5	91	10	Q9SDP4	Q9SDP4 brassica ca
741	1.5	63	16	Q8EJMI	Q8EJMI escherichia	814	6	1.5	91	13	Q8QGS6	Q8QGS6 gallus gall
742	1.5	63	17	Q9HZZ1	Q9HZZ1 halobacteri	815	6	1.5	91	16	Q99QC2	Q99QC2 streptomyce
743	1.5	64	16	Q8RFJ2	Q8RFJ2 fusobacteri	816	6	1.5	91	16	Q8EDT2	Q8EDT2 shewanella
744	1.5	65	16	Q92ZEL	Q92ZEL rhizobium m	817	6	1.5	92	3	Q59914	Q59914 neotyphodiu
745	1.5	66	16	Q8NTK3	Q8NTK3 corynebacte	818	6	1.5	92	3	Q59910	Q59910 balansta sp
746	1.5	67	4	Q8NE92	Q8NE92 homo sapien	819	6	1.5	92	3	Q59913	Q59913 neotyphodiu

820	6	1.5	92	16	P95120	P95120 mycobacteri	833	6	1.5	102	16	050482	050482 streptomyc
821	6	1.5	92	16	Q8XBH5	Q8XBH5 escherichia	834	6	1.5	103	5	009209	009209 agnus conv
822	6	1.5	92	16	Q9XBE1	Q9XBE1 streptomyc	835	6	1.5	103	6	09XST6	09XST6 caris faml
823	6	1.5	92	16	Q8ETU6	Q8ETU6 oceanobacil	836	6	1.5	103	16	08XX10	08XX10 ralsstonia s
824	6	1.5	93	2	Q9Z915	Q9Z915 escherichia	837	6	1.5	103	16	0920P9	0920P9 rhizobium m
825	6	1.5	93	16	Q9ZHL7	Q9ZHL7 rickettsia	838	6	1.5	104	16	091LR7	091LR7 arabidopsis
826	6	1.5	94	10	Q9SDU0	Q9SDU0 brassica na	839	6	1.5	104	10	08GSD8	08GSD8 oryza sativ
827	6	1.5	94	16	Q92KE9	Q92KE9 rhizobium m	890	6	1.5	104	16	P73134	P73134 synechocyst
828	6	1.5	95	2	Q48906	Q48906 mycobacteri	901	6	1.5	104	16	09HTW3	09HTW3 pseudomonas
829	6	1.5	95	2	Q50749	Q50749 mycobacteri	902	6	1.5	104	16	098NA6	098NA6 rhizobium l
830	6	1.5	95	2	Q57364	Q57364 mycobacteri	903	6	1.5	104	16	Q983R7	Q983R7 rhizobium l
831	6	1.5	95	10	Q9SP96	Q9SP96 actinidia d	904	6	1.5	105	5	Q25305	Q25305 leishmania
832	6	1.5	95	16	Q93R20	Q93R20 streptomyc	905	6	1.5	105	5	Q9TVI8	Q9TVI8 leishmania
833	6	1.5	95	16	Q8DL04	Q8DL04 synechococc	906	6	1.5	105	16	Q9A4R3	Q9A4R3 calrobacter
834	6	1.5	96	2	Q9S552	Q9S552 borrelia bu	907	6	1.5	105	16	08XVC1	08XVC1 ralsstonia s
835	6	1.5	96	2	Q9K534	Q9K534 listeria mo	908	6	1.5	105	16	08FAH2	08FAH2 leptospira
836	6	1.5	96	5	Q9BMW8	Q9BMW8 leishmania	909	6	1.5	105	16	Q9EB76	Q9EB76 shewanella
837	6	1.5	96	16	Q9CDE1	Q9CDE1 mycobacteri	910	6	1.5	105	17	Q9YE44	Q9YE44 aeropyrum p
838	6	1.5	97	12	Q91HD4	Q91HD4 human polio	911	6	1.5	106	6	08HY29	08HY29 ovis aries
839	6	1.5	97	12	Q91HG0	Q91HG0 human polio	912	6	1.5	106	15	Q41402	Q41402 human immun
840	6	1.5	97	12	Q91HF4	Q91HF4 human polio	913	6	1.5	106	16	Q91630	Q91630 pseudomonas
841	6	1.5	97	12	Q91HF5	Q91HF5 human polio	914	6	1.5	106	17	08ZTC9	08ZTC9 pyrobaculum
842	6	1.5	97	12	Q91HE4	Q91HE4 human polio	915	6	1.5	107	5	08SRP1	08SRP1 encephalito
843	6	1.5	97	12	Q91HE2	Q91HE2 human polio	916	6	1.5	107	16	Q9UYR2	Q9UYR2 neisseria m
844	6	1.5	97	12	Q91HE1	Q91HE1 human polio	917	6	1.5	107	16	08YXE0	08YXE0 anabaena sp
845	6	1.5	97	12	Q91HG1	Q91HG1 human polio	918	6	1.5	107	16	08YI88	08YI88 bruceella me
846	6	1.5	97	12	Q91HD7	Q91HD7 human polio	919	6	1.5	107	16	08YIW8	08YIW8 ralsstonia s
847	6	1.5	97	12	Q91HE5	Q91HE5 human polio	920	6	1.5	107	16	08FZM8	08FZM8 bruceella su
848	6	1.5	97	12	Q91HE7	Q91HE7 human polio	921	6	1.5	108	3	Q94724	Q94724 schizosacch
849	6	1.5	97	12	Q91HE3	Q91HE3 human polio	922	6	1.5	108	5	Q26290	Q26290 aedes aegy
850	6	1.5	97	12	Q91HE1	Q91HE1 human polio	923	6	1.5	108	13	Q8AYB4	Q8AYB4 salvelinus
851	6	1.5	97	12	Q91HD8	Q91HD8 human polio	924	6	1.5	108	16	Q92W48	Q92W48 rhizobium m
852	6	1.5	97	12	Q91HE9	Q91HE9 human polio	925	6	1.5	108	17	Q9Y922	Q9Y922 aeropyrum p
853	6	1.5	97	12	Q91HE0	Q91HE0 human polio	926	6	1.5	109	2	Q9AJH0	Q9AJH0 vibrio prot
854	6	1.5	97	12	Q91HE8	Q91HE8 human polio	927	6	1.5	109	2	Q9ANF1	Q9ANF1 bradyrhizob
855	6	1.5	97	12	Q91HE7	Q91HE7 human polio	928	6	1.5	109	6	Q29286	Q29286 sus scrofa
856	6	1.5	97	12	Q91HD3	Q91HD3 human polio	929	6	1.5	109	9	Q94MQ9	Q94MQ9 bacterioph
857	6	1.5	97	12	Q91HD6	Q91HD6 human polio	930	6	1.5	109	16	Q25421	Q25421 helicobacte
858	6	1.5	97	12	Q91HP9	Q91HP9 human polio	931	6	1.5	109	16	Q8NMV2	Q8NMV2 cornebacte
859	6	1.5	97	12	Q91HP8	Q91HP8 human polio	932	6	1.5	109	16	Q9X948	Q9X948 streptomyc
860	6	1.5	97	12	Q91HE3	Q91HE3 human polio	933	6	1.5	110	2	Q54878	Q54878 streptococ
861	6	1.5	97	12	Q91HE6	Q91HE6 human polio	934	6	1.5	110	2	Q33004	Q33004 mycobacteri
862	6	1.5	97	12	Q91HE0	Q91HE0 human polio	935	6	1.5	110	5	Q8ISN2	Q8ISN2 myrmecia ni
863	6	1.5	97	12	Q91HD5	Q91HD5 human polio	936	6	1.5	110	12	Q8OW60	Q8OW60 norwalk-lik
864	6	1.5	97	12	Q91HE2	Q91HE2 human polio	937	6	1.5	110	16	Q8YOE4	Q8YOE4 ralsstonia s
865	6	1.5	97	12	Q91HD9	Q91HD9 human polio	938	6	1.5	110	16	Q8PER0	Q8PER0 xanthomonas
866	6	1.5	97	12	Q91HE6	Q91HE6 human polio	939	6	1.5	110	17	Q8ZSN0	Q8ZSN0 pyrobaculum
867	6	1.5	98	12	P89937	P89937 gbv-a-lik	940	6	1.5	111	10	Q94ZU5	Q94ZU5 oryza sativ
868	6	1.5	98	12	P90501	P90501 unidentified	941	6	1.5	112	2	Q91BE9	Q91BE9 helicobacte
869	6	1.5	98	16	Q92T89	Q92T89 rhizobium m	942	6	1.5	112	10	Q814D1	Q814D1 oryza sativ
870	6	1.5	98	16	Q8RD90	Q8RD90 thermomae	943	6	1.5	112	11	Q9QY78	Q9QY78 raltus norv
871	6	1.5	98	16	Q8RF35	Q8RF35 leptospira	944	6	1.5	112	16	Q92K70	Q92K70 rhizobium m
872	6	1.5	99	16	Q8F3E5	Q8F3E5 oryza sativ	945	6	1.5	112	16	Q06774	Q06774 mycobacteri
873	6	1.5	99	13	Q8AYJ2	Q8AYJ2 squalus aca	946	6	1.5	113	2	Q9X9U7	Q9X9U7 streptomyc
874	6	1.5	99	15	Q91LP2	Q91LP2 human immun	947	6	1.5	113	16	Q98XK14	Q98XK14 rhizobium l
875	6	1.5	99	16	Q8ECT8	Q8ECT8 shewanella	948	6	1.5	113	16	Q8XCT2	Q8XCT2 escherichia
876	6	1.5	100	2	Q50146	Q50146 mycobacteri	949	6	1.5	113	16	Q8ZSM0	Q8ZSM0 yersinia pe
877	6	1.5	100	2	Q9FP94	Q9FP94 rhodobacter	950	6	1.5	113	16	Q8CVZ6	Q8CVZ6 escherichia
878	6	1.5	100	8	Q32758	Q32758 oryza sativ	951	6	1.5	114	2	Q8VL88	Q8VL88 bacillus su
879	6	1.5	100	10	Q9FUZ8	Q9FUZ8 arabidopsis	952	6	1.5	114	2	Q8VOA5	Q8VOA5 bacillus su
880	6	1.5	100	12	Q9WB51	Q9WB51 hepatitis c	953	6	1.5	114	2	Q8VOA7	Q8VOA7 bacillus su
881	6	1.5	100	18	Q9DUC2	Q9DUC2 tt virus. o	954	6	1.5	114	2	Q8VOA4	Q8VOA4 bacillus su
882	6	1.5	101	5	Q9NKO9	Q9NKO9 cypanosoma	955	6	1.5	114	8	Q9MFE4	Q9MFE4 beta vulgar
883	6	1.5	101	6	Q9XSO2	Q9XSO2 ovis aries	956	6	1.5	114	11	Q91YF3	Q91YF3 mus musculu
884	6	1.5	101	16	Q9CDX3	Q9CDX3 lactococcus	957	6	1.5	114	12	Q91YF3	Q91YF3 mus musculu
885	6	1.5	101	16	Q8EIM9	Q8EIM9 shewanella	958	6	1.5	114	13	Q9DFP8	Q9DFP8 oncorhynch
886	6	1.5	101	16	Q8E7T0	Q8E7T0 streptococ	959	6	1.5	115	2	Q51045	Q51045 neisseria g
887	6	1.5	101	16	Q8E5G8	Q8E5G8 streptococ	960	6	1.5	115	2	Q51053	Q51053 rhodococcus
888	6	1.5	101	16	Q8E2C3	Q8E2C3 streptococ	961	6	1.5	115	9	Q9AZA5	Q9AZA5 lactobacilli
889	6	1.5	101	16	Q8DZR9	Q8DZR9 streptococ	962	6	1.5	115	10	Q8Z787	Q8Z787 oryza sativ
890	6	1.5	101	16	Q8DS24	Q8DS24 streptococ	963	6	1.5	115	10	Q96N09	Q96N09 oryza sativ
891	6	1.5	102	10	Q8S4Q9	Q8S4Q9 cryptocodi	964	6	1.5	115	16	Q9RXJ1	Q9RXJ1 delnecoccus
892	6	1.5	102	11	Q99KG2	Q99KG2 mus musculu	965	6	1.5	115	16	Q8P4B2	Q8P4B2 xanthomonas

966	6	1.5	115	16	Q91210	O91210 streptomyc
967	6	1.5	116	2	O51016	O51016 neisseria g
968	6	1.5	116	2	O8GB96	O8GB96 heliobacill
969	6	1.5	116	10	Q9SNR0	Q9SNR0 oryza sativ
970	6	1.5	116	10	Q9SNR7	Q9SNR7 oryza sativ
971	6	1.5	116	10	Q9ZTR3	Q9ZTR3 physalis ci
972	6	1.5	116	10	O8GZB0	O8GZB0 triticum ae
973	6	1.5	116	16	O9RV55	O9RV55 delinococcus
974	6	1.5	116	16	O9X9T5	O9X9T5 streptomyces
975	6	1.5	116	17	Q9H0G9	Q9H0G9 halobacteri
976	6	1.5	117	2	O51017	O51017 neisseria g
977	6	1.5	117	2	Q9WTG4	Q9WTG4 escherichia
978	6	1.5	117	16	Q9ZWZ6	Q9ZWZ6 rhizobium m
979	6	1.5	117	16	O8PFW6	O8PFW6 xanthomonas
980	6	1.5	117	16	O9RDE3	O9RDE3 streptomyces
981	6	1.5	118	2	Q93HL1	Q93HL1 streptomyces
982	6	1.5	118	10	O8S745	O8S745 oryza sativ
983	6	1.5	118	12	O11279	O11279 human papil
984	6	1.5	118	16	Q9A7X0	Q9A7X0 caulobacter
985	6	1.5	118	16	O8Y1U5	O8Y1U5 ralslonia s
986	6	1.5	119	2	O9AN60	O9AN60 bradyrhizob
987	6	1.5	119	2	O52505	O52505 pseudomonas
988	6	1.5	119	16	Q9PFZ4	Q9PFZ4 xyella fas
989	6	1.5	119	16	O8ZKK2	O8ZKK2 salmonella
990	6	1.5	119	16	O97GR3	O97GR3 clostridium
991	6	1.5	119	16	O8DJ95	O8DJ95 synechococc
992	6	1.5	120	2	P74898	P74898 thermus the
993	6	1.5	120	10	O8LHY6	O8LHY6 oryza sativ
994	6	1.5	121	3	O94217	O94217 cilioporiu
995	6	1.5	121	8	O9TG24	O9TG24 rhinopolus
996	6	1.5	121	9	O80111	O80111 bacterioph
997	6	1.5	121	10	O8GW10	O8GW10 arabidopsis
998	6	1.5	121	16	O9KAG1	O9KAG1 bacillus ha
999	6	1.5	121	16	O8Y357	O8Y357 ralslonia s
1000	6	1.5	121	17	Q96ZP1	Q96ZP1 sulfobobus

ALIGNMENTS

RESULT 1

Q9JTS0	PRELIMINARY;	PRT;	412 AA.
AC	O9JTS0;		
AD	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)		
DE	Membrane fusion protein.		
GN	MTFC OR NMA1970.		
OS	Neisseria meningitidis (serogroup A).		
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;		
OC	Neisseriaceae; Neisseria.		
OK	NCBI_TaxID=65699;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Z2491 / Serogroup A / Serotype 4A;		
RX	MEDLINE=2022556; PubMed=10761919;		
RA	Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,		
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,		
RA	Javels R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,		
RA	Jagals K., Leather S., Moulé S., Mungall K., Quail M.A.,		
RA	Rajandream K.M., Rutherford K.M., Simmonds M., Skelton J.,		
RA	Whitehead S., Spratt B.G., Barrall B.G.;		
RT	"Complete DNA sequence of a serogroup A strain of Neisseria		
RT	meningitidis Z2491."		
RL	Nature 404:502-506(2000).		
DR	EMBL; AL162757; CAB85190.1; -		
DR	InterPro; IPR006143; H1YD.		
DR	Pfam; PF00529; H1YD; 1.		
KW	Complete proteome.		
SO	SEQUENCE 412 AA; 42954 MW; 5CF797BA370AA75D CRC64;		

Query Match 41.5%; Score 171; DB 16; Length 412;

Qy	1	MAFAFRMRAAALAAVAATLVS	CGCGDAAOGGAPAGRAPAPVYGVVTHVHTVLT	60
Db	1	MAFAFRMRAAALAAVAATLVS	CGCGDAAOGGAPAGRAPAPVYGVVTHVHTVLT	60
Qy	61	VELGRLESRTADVRAOVGGII	IOKRLFOEGSVYRACOPLYQIISSTYENALLESAPQLA	120
Db	61	VELGRLESRTADVRAOVGGII	IOKRLFOEGSVYRACOPLYQIISSTYENALLESAPQLA	120
Qy	121	TAQATLAKADADLARYKPLVA	AEVSRQEDAAVTAKRSAPGYKAAQAAIKSAGINLR	180
Db	121	TAQATLAKADADLARYKPLVA	AEVSRQEDAAVTAKRSAPGYKAAQAAIKSAGINLR	180
Qy	181	SRITAPISGFTGQSKVSEGL	TLNAGDTVLTATIQTNPMYVNTQSASEVKKLRQIAEG	240
Db	181	SRITAPISGFTGQSKVSEGL	TLNAGDTVLTATIQTNPMYVNTQSASEVKKLRQIAEG	240
Qy	241	KLLAADGVIAVGIRFDDGTV	PEKGRLLFADP 272	
Db	241	KLLAADGVIAVGIRFDDGTV	PEKGRLLFADP 272	

RESULT 2

Q9JY66	PRELIMINARY;	PRT;	412 AA.
ID	O9JY66;		
AC	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)		
DE	Membrane fusion protein.		
GN	NMB1716.		
OS	Neisseria meningitidis (serogroup B).		
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;		
OC	Neisseriaceae; Neisseria.		
OK	NCBI_TaxID=491;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MC58 / Serogroup B;		
RX	MEDLINE=20175755; PubMed=10710307;		
RA	Tetzelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,		
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,		
RA	Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.R.,		
RA	Haff D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,		
RA	Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,		
RA	Cotton M.D., Ullerbach T.R., Khouri H., Qin H., Yamachyan J.,		
RA	Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,		
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;		
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain		
RT	MC58."		
RL	Science 287:1809-1815(2000).		
DR	EMBL; AE002521; AAF42063.1; -		
DR	TIGR; NMB1716; -		
DR	InterPro; IPR006143; H1YD.		
DR	Pfam; PF00529; H1YD; 1.		
KW	Complete proteome.		
SO	SEQUENCE 412 AA; 42795 MW; 087A0DB0E3113E4E CRC64;		

Query Match 32.3%; Score 133; DB 16; Length 412;
 Best Local Similarity 100.0%; Pred. No. 2e-115;
 Matches 133; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	274	VNESTGQITLRAAVNDONITL	MPGLVYVLMDOVAVDNAFVVPQAVTRGAKDTVMIVNA	333
Db	274	VNESTGQITLRAAVNDONITL	MPGLVYVLMDOVAVDNAFVVPQAVTRGAKDTVMIVNA	333
Qy	334	OGMEPREVTAAGQGTMIWTS	GLKDGKVVVEGISIAGITGAKVTPKEMASSENOAA	393
Db	334	OGMEPREVTAAGQGTMIWTS	GLKDGKVVVEGISIAGITGAKVTPKEMASSENOAA	393
Qy	394	APQSGVOTASEAK 406		
Db	394	APQSGVOTASEAK 406		

Db 394 APOGVCQTASEAK 406

RESULT 3

ID 051007 PRELIMINARY; PRT; 271 AA.

AC 051007; PRELIMINARY; PRT; 271 AA.

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Mtrc protein (Fragment).

GN MTRC.

OS Neisseria gonorrhoeae.

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OC Neisseriaceae; Neisseria.

OX NCBI_TaxID=485;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CH95;

RX MEDLINE=94254732; PubMed=8196548;

RA Pan W., Spratt B.G.;

RT "Regulation of the permeability of the gonococcal cell envelope by the mtr system.";

RL Mol. Microbiol. 11:769-775(1994).

DR EMBL; 225796; CAA81046.1; -.

DR InterPro; IPR006143; HLYD.

DR Pfam; PF00529; HLYD; 1.

FT NON TER 271

FT SEQUENCE 271 AA; 28145 MW; 1DDADD1F335B39C5 CRC64;

Query Match 24.3%; Score 100; DB 2; Length 271;

Best Local Similarity 100.0%; Pred. No. 9,1e-85;

Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 OEYDAAYAKSAEAGVKAQAQAIKAGININRSRTIPISGFIQSKVSEGTLLNAGDT 207

Db 148 OEYDAAYAKSAEAGVKAQAQAIKAGININRSRTIPISGFIQSKVSEGTLLNAGDT 207

QY 208 TVLATIRGTNPMYVAVTOSASEVMTLRQIAEGKLIADG 247

Db 208 TVLATIRGTNPMYVAVTOSASEVMTLRQIAEGKLIADG 247

RESULT 4

ID 08X7E1 PRELIMINARY; PRT; 373 AA.

AC 08X7E1; PRELIMINARY; PRT; 373 AA.

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Putative efflux pump (Putative multidrug-efflux transport protein).

GN Z2509 OR ECS1863.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoculis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RT Nature 409:529-533(2001).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

DR EMBL; AE005377; AAG56521.1; -.

DR EMBL; AP002556; BAB35286.1; -.

DR InterPro; IPR006143; HLYD.

DR Pfam; PF00529; HLYD; 1.

KM Complete proteome.

SQ SEQUENCE 373 AA; 39612 MW; 2A573537AB4CDBF CRC64;

Query Match 2.7%; Score 11; DB 16; Length 373;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 QVGGIIQKRLP 88

Db 63 QVGGIIQKRLP 73

RESULT 5

ID 08GC84 PRELIMINARY; PRT; 374 AA.

AC 08GC84; PRELIMINARY; PRT; 374 AA.

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Befa lipoprotein precursor.

GN EEFA.

OS Enterobacter aerogenes (Aerobacter aerogenes).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Enterobacter.

OX NCBI_TaxID=548;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BW16627;

RA Maai M., Pages J.M., Pradel E.;

RT "Identification and characterization of the Enterobacter aerogenes RT eefABC operon encoding a putative tripartite efflux system.";

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ508047; CAD4861.1; -.

KW Signal; Lipoprotein.

FT SIGNAL 1 17

FT CHAIN 18 374

FT POTENTIAL.

FT EEFA LIPOPROTEIN.

SQ SEQUENCE 374 AA; 39374 MW; 332BFD6FAB942BB CRC64;

Query Match 2.7%; Score 11; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 QVGGIIQKRLP 88

Db 64 QVGGIIQKRLP 74

RESULT 6

ID 08CM42 PRELIMINARY; PRT; 384 AA.

AC 08CM42; PRELIMINARY; PRT; 384 AA.

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Acriflavine resistance protein A precursor.

GN C1761.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=06.H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Weich R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blatcher F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 KW EMBL; AF016760; AAN80227.1; -.
 SQ Complete proteome.
 SO SEQUENCE 384 AA; 41129 MW; FC9E69E4583EC7B CRC64;

Query Match 2.7%; Score 11; DB 16; Length 384;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 78 QVGGIIQKRLF 88
 |||||
 DB 74 QVGGIIQKRLF 84

RESULT 7
 O9AA03 PRELIMINARY; PRT; 388 AA.

AC O9AA03; 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE HlyD family secretion protein.
 GN CC0808.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 CC Caulobacteriaceae; Caulobacter.
 OX NCBI_TaxID=155892;
 (1)
 RP SEQUENCE FROM N.A.
 RN RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Fejdylyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin M.B., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
 DR EMBL; AB005757; AAK22793.1; -.
 DR TIGR; CC0808; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 KW Complete proteome.
 SQ SEQUENCE 388 AA; 39483 MW; 4D934C3F69052DAD CRC64;

Query Match 2.7%; Score 11; DB 16; Length 388;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 VRAAGPLYQID 104
 |||||
 DB 89 VRAAGPLYQID 99

RESULT 8
 O969P9 PRELIMINARY; PRT; 124 AA.

AC O969P9; 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE CKII beta binding protein 2.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Muscle;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yoon S.-H., Bae Y.-S.;
 RT "Homo sapiens CXBB2 mRNA.";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC013039; AAH13039.1; -.
 DR EMBL; BC013027; AAH13027.1; -.
 DR EMBL; AF412816; AAL55291.1; -.
 SQ SEQUENCE 124 AA; 14494 MW; 9C50C85B061DFD4A CRC64;

Query Match 2.4%; Score 10; DB 4; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.51;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
 |||||
 DB 104 RAALAAVA 113

RESULT 9
 O81VM3 PRELIMINARY; PRT; 216 AA.

AC O81VM3; 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein (fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Revolletta C.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Frigimelica E., Lanfranchi G.;
 RT "Full length sequencing of some human and murine muscular transcripts
 (Telethon Italy Project B41).";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ295986; CAC82499.1; -.
 KW Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 216 AA; 24768 MW; D7DE609D6188B835 CRC64;

Query Match 2.4%; Score 10; DB 4; Length 216;
 Best Local Similarity 100.0%; Pred. No. 0.83;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
 |||||
 DB 196 RAALAAVA 205

RESULT 10
 O9BSM6 PRELIMINARY; PRT; 218 AA.

AC O9BSM6; 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 2310040617 gene (fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=skin;
RA Strusberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004944; AA04944.1; -.
FT NON TER
SQ SEQUENCE 218 AA; 25023 MW; 71BA2752F4CFEDB CRC64;

Query Match
Best Local Similarity 100.0%; Score 10; DB 4; Length 218;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
Db 198 RAALAAVA 207

RESULT 11
OBTES1 PRELIMINARY; PRT; 222 AA.
AC OBTES1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Papillomavirus L2 interacting nuclear protein 1.
GN PLINP-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mueller M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Goernemann J.;
RL "Cellular Interaction Partners of the HPV Minor Capsid Protein L2.";
RL Thesis (2002), Department of Fakultät fuer Biologie,
RL Universitaet Heidelberg, Heidelberg, Germany.
DR EMBL; AJ429498; CAD2344.1; -.
KW Nuclear protein.
SQ SEQUENCE 222 AA; 25368 MW; 21EF67045B1F5868 CRC64;

Query Match
Best Local Similarity 100.0%; Score 10; DB 4; Length 222;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
Db 202 RAALAAVA 211

RESULT 12
OBTAB8 PRELIMINARY; PRT; 222 AA.
AC OBTAB8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CR6 interacting factor 1 (CKII beta associating protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung H., Yi Y., Jung N.-C., Kim D., Shong M.;
RL "CR6 interacting factor 1 (CRIF1), a novel nuclear protein that
interacts with Gadd45 family proteins. Cloning, functional

```

```

RT characterization and analysis of expression.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yoon S.H., Bae Y.S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF479749; AL65877.1; -.
DR EMBL; AF475095; AAM10639.1; -.
SQ SEQUENCE 222 AA; 25384 MW; 304C397AEC958804 CRC64;

Query Match
Best Local Similarity 100.0%; Score 10; DB 4; Length 222;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAVA 19
Db 202 RAALAAVA 211

RESULT 13
OBTZW1 PRELIMINARY; PRT; 444 AA.
AC OBTZW1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 23, Last annotation update)
DE Putative HlyD family secretion protein (Putative membrane
protein).
GN IPO2847 OR Y1386.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagsels K., Katlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G., III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.V., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
RN [3]
RA EMBL; AJ414154; CAC92099.1; -.
DR EMBL; AE013741; AAM84958.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KW Complete proteome.
SQ SEQUENCE 444 AA; 47514 MW; BF70B791FF55BD01 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 444;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 181 SRTAPISG 189
Db 198 SRTAPISG 206

```

```

RESULT 14
ID 09J007 PRELIMINARY; PRT; 107 AA.
AC 09J007;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
GN 50S ribosomal protein L24.
OR RPLX OR NMA0118 OR NMB0153.
OS Neisseria meningitidis (serogroup A), and
OC Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jags R., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature. 404:502-506 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC68 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Hafz D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cline H., Clark E.B.,
RA Cotton M.D., Utechtack T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC68."
RL Science. 287:1809-1815 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AL162752; PubMed=10710307;
RX EMBL; AL162752; CAB83433.1; -;
DR EMBL; AE002373; AAF40611.1; -;
DR TIGR; NMB0153; -;
DR InterPro; IPR005824; KOW; 1.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; RIBosomal_L24.
DR InterPro; IPR005825; RIBosomal_L24_26.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; RIBosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGRfam; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
DR RIBosomal protein; Complete proteome.
SQ SEQUENCE 107 AA; 11667 MW; EICEAA2FOA93A24C CRC64;

Query Match 1.9%; Score 8; DB 16; Length 107;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DE P0712E02.9 protein (OSUNB0024F06.19 protein).
GN P0712E02.9 OR OSUNB0024F06.19.
OS Oryza sativa (Rice), and
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530, 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=O. sativa; STRAIN=cv. Nipponbare;
RA Sasaki T., Matsunoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (Gh3) genomic DNA, chromosome 1, PAC
RT clone:P0712E02."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=O. sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
RA Sasaki T., Matsunoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT clone:OSUNB0024F06."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003492; BAB1846.1; -;
DR EMBL; AP003341; BAC01241.1; -;
DR Gramene; Q94D46; -;
SQ SEQUENCE 125 AA; 12949 MW; 0488F2AB81F7C6C0 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 125;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 16
ID 032184 PRELIMINARY; PRT; 129 AA.
AC 032184;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
GN YUSR protein.
OR YUSR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolochin A., Borchert S.,
RA Borstis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle B., Rapoport G., Rey M., Reynolds S.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

```

RA Setaiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanaka K., Tanaka T., Terpstra P., Tognoni A.,
 RA Toesato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zimstein B., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC EMBL: 299120; CAB15279.1; -.
 DR HSSP; 070351; 186W.
 DR InterPro; IPR002198; ADH_short.
 DR Pfam; PF00106; adh_short; 1.
 DR PRINTS; PR00080; SDRFAMILY.
 DR PROSITE; PS00061; ADH_SHORT; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 129 AA; 13477 MW; 1874C97295B292A CRC64;

Query Match 1.9%; Score 8; DB 16; Length 129;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 370 SIAGITGA 377
 Db 28 SIAGITGA 35

RESULT 17
 O9RPUI PRELIMINARY; PRT; 156 AA.
 AC O9RPUI;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE Anaerobically induced outer membrane protein Alpi (Fragment).
 GN Alpi.
 OS *Salmonella typhi*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Salmonella*.
 NC NCB1_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TY2;
 RA Maulen N.P., Mora G.C.;
 RT "An anaerobically induced outer membrane protein of *Salmonella typhi*
 RT is necessary for the efficient invasion of Hep-2 cells."
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF145591; AAF08700.1; -.
 FT NON TER 1 156
 FT 156
 SQ SEQUENCE 156 AA; 16591 MW; 1EC862A52419BCC6 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 156;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 RAALAAA 17
 Db 13 RAALAAA 20

RESULT 18
 O8SOKI PRELIMINARY; PRT; 162 AA.
 AC O8SOKI;

DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE B1078G07.34 protein (P0696G06.11 protein).
 GN B1078G07.34 OR P0696G06.11
 OS *Oryza sativa* (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; *Oryza*.
 NC NCB1_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
 RT clone: B1078G07.";
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone: P0696G06.";
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AP003407; BAB90212.1; -.
 DR EMBL; AP003316; BAC06254.1; -.
 KW Gramine; O8SOKI; -.
 SQ SEQUENCE 162 AA; 18253 MW; 873B89784D3BDBF4 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 162;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 15 AAAVALVT 22
 Db 8 AAAVALVT 15

RESULT 19
 O8XCR9 PRELIMINARY; PRT; 171 AA.
 AC O8XCR9;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative fibrial protein (Putative fibrial-like protein).
 GN SPMF OR 20691 OR ECS0596.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 NC NCB1_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7,"
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).
 DR EMBL; AB005234; AAG54891.1; -
 DR EMBL; AP002552; BAB34019.1; -
 DR InterPro: IPR002559; Fimbrilal.
 DR Pfam; PF00419; Fimbrilal; 1.
 KW Complete proteome.
 SQ SEQUENCE 171 AA; 18212 MW; 629C9DC02FC83516 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 171;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 LFNAGDTT 208
 Db 61 LFNAGDTT 68

RESULT 20

08XC98 PRELIMINARY; PRT; 183 AA.
 AC 08XC98;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DE Hypothetical protein z2047.
 GN 22047 OR ECS1767.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OK NCBI_TaxId=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Požar G., Hackett J., Klink S., Boutin A., Shao Y., Miller B.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.,
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.,
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AB005343; AAG56122.1; -
 DR EMBL; AP002556; BAB35190.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 183 AA; 21220 MW; 5A6C3E3B9000796E CRC64;

Query Match 1.9%; Score 8; DB 16; Length 183;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 LFNADPVV 274
 Db 130 LFNADPVV 137

RESULT 21

09YAH2 PRELIMINARY; PRT; 183 AA.
 AC 09YAH2;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Hypothetical protein ABE1967.
 GN ABE1967.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcates;
 CC Desulfurococcaceae; Aeropyrum.
 OK NCBI_TaxId=56636;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=993310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Akai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000063; BAA80977.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 183 AA; 19724 MW; 604085540DBA0ABC CRC64;

Query Match 1.9%; Score 8; DB 17; Length 183;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ALAAVAL 20
 Db 118 ALAAVAL 125

RESULT 22

08P176 PRELIMINARY; PRT; 206 AA.
 AC 08P176;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DE Hypothetical protein XAC3022.
 GN XAC3022.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xanthomonas.
 OK NCBI_TaxId=92829;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferrero J.A., Reinach P.C., Farah C.S., Furlan L.R.,
 RA Alvares L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cammaral F., Cardoso J., Chambez F., Ciapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferrero M.I.T.,
 RA Forniglieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
 RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.,
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RL Nature 417:459-463(2002).
 DR EMBL; AB011943; AAM37867.1; -
 DR InterPro: IPR00104; Antifreeze_1.
 DR PRINTS; PR00308; ANTIFREEZE1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 206 AA; 20298 MW; 88C31E3F624A908 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 206;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAAA 17
 |||||
 Db 32 RAALAAAA 39

RESULT 23

O9KZV1 PRELIMINARY; PRT; 206 AA.

AC O9KZV1
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Putative ABC transport system transmembrane protein.
 GN SC04150 OR SCD84.17.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomycetes.
 CX NCBI_TaxID=1902;

SEQUENCE FROM N.A.
 [1]
 RP STRAIN=A3(2) / M145;
 RC MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James A.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL: AL939119; CAB88483.1; -.
 KW Transmembrane; Complete proteome.
 SQ SEQUENCE 206 AA; 21043 MW; A3D81PAD92828A39 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 206;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 LAAAVAV 21
 |||||
 Db 184 LAAAVAV 191

RESULT 24

P95115 PRELIMINARY; PRT; 214 AA.

AC P95115
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein RV2980.
 GN RV2980 OR MT3058 OR MTCY349.07C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1773;

SEQUENCE FROM N.A.
 [1]
 RP STRAIN=H37RV;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigemeier K., Gao S., Barry C.E. III, Tebala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 [2]

SEQUENCE FROM N.A.
 [2]
 RP STRAIN=CDC 1551 / Oshkosh;
 RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Hatt D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z83018; CAB05432.1; ALT INIT.
 DR EMBL: AE007126; AAK47385.1; -.
 DR TIGR: MT3058; -.
 DR TubercuList; RV2980; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 214 AA; 22522 MW; EDD46F75C351D96E CRC64;

Query Match 1.9%; Score 8; DB 16; Length 214;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALAAAV 18
 |||||
 Db 49 AALAAAV 56

RESULT 25

O8SPE7 PRELIMINARY; PRT; 222 AA.

AC O8SPE7
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative papillomavirus L2 interacting nuclear protein.
 GN PLINP.
 OS Cercopithecus aethiops (Green monkey) (Giveti).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Cercopithecus.
 CX NCBI_TaxID=9534;

SEQUENCE FROM N.A.
 [1]
 RP Mueller M.;
 RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Goernemann J.;
 RT "Cellular interaction Partners of the HPV Minor Capsid Protein L2.";
 RL Thesis (2002), Department of Fakultät fuer Biologie,
 RL Universitaet Heidelberg, Heidelberg, Germany.
 DR EMBL: AJ437508; CAD26811.1; -.
 KW Nuclear protein.
 SQ SEQUENCE 222 AA; 25378 MW; 1DEDF23D38DAE04 CRC64;

Query Match 1.9%; Score 8; DB 6; Length 222;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAAA 17
 |||||
 Db 202 RAALAAAA 209

RESULT 26

O9Z5W3

ID Q925W3 PRELIMINARY; PRT; 238 AA.
 AC Q925W3;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative transcriptional regulatory protein Ohbr.
 GN OHBR.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OK NCBI_TaxId=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=142;
 RX MEDLINE=99240444; PubMed=10224014;
 RA Tsoi T.V., Plotnikova E.G., Cole J.R., Guerin W.F., Bagdasarian M.,
 Tiedje J.M.,
 "Cloning, expression, and nucleotide sequence of the Pseudomonas
 aeruginosa 142 ohb genes coding for oxygenolytic ortho dehalogenation
 of halobenzoates.";
 RT Appl. Environ. Microbiol. 65:2151-2162 (1999).
 CC -1- SIMILARITY: BELONGS TO THE ICRL FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; AF121970; AAD20004.1;
 DR InterPro; IPR005471; HTH_ICLR.
 DR Pfam; PF01614; ICLR; 1.
 DR DNA-binding; Transcription; Transcription regulation.
 SQ SEQUENCE 238 AA; 25692 MW; 59CG78660D28A67E CRC64;

Query Match 1.9%; Score 8; DB 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 KLRQIAE 239
 |||||
 DB 152 KLRQIAE 159

RESULT 27
 ID Q96B43 PRELIMINARY; PRT; 253 AA.
 AC Q96B43;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Similar to methionyl aminopeptidase 2 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis, and Embryonic carcinoma;
 RA Straussberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC016008; AAH16008.1; -
 DR InterPro; IPR002468; MAP_2.
 DR InterPro; IPR001714; Methamino_Pase.
 DR InterPro; IPR000994; Peptidase_M24.
 DR Pfam; PF00557; Peptidase_M24; 1.
 DR PRINTS; PRO0599; MAPPEPTIDASE.
 DR TRGFAMS; TIGR00501; met_pdae_II; 1.
 DR PROSITE; PS01202; MAP_2; 1.
 KW Aminopeptidase.
 FT NON_TER 1
 SQ SEQUENCE 253 AA; 28230 MW; B860E2D127731D6 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 253;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||

DB 10 NAGDTTVL 17
 RESULT 28
 ID Q9S295 PRELIMINARY; PRT; 290 AA.
 AC Q9S295;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative secreted lipase.
 GN SC01735 OR SC111.24C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
 OK NCBI_TaxId=1902;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D., Harris D.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96 (1996).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147 (2002).
 DR EMBL; AL939110; CAB50950.1; -
 DR HSSP; Q05489; ITH.
 DR InterPro; IPR002918; Lipase_2.
 DR InterPro; IPR000379; Ser_estr_site.
 DR Pfam; PF01674; Lipase_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 290 AA; 30496 MW; D56E8508350F7432 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 290;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 14 LAAVAALV 21
 |||||
 DB 16 LAAVAALV 23

RESULT 29
 ID Q926G6 PRELIMINARY; PRT; 296 AA.
 AC Q926G6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Putative sugar uptake ABC transporter periplasmic solute-binding
 DE protein.
 GN RB0955 OR SMB2137.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC Plasmid pSymB (megaplasmid 2).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
 RA Golding B., Pecher A.,
 RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL; AL603645; CAC49355.1; -
 DR InterPro: IPR001761; PeriplasmBP/Lact.
 DR Pfam: PF00532; Peripla_BP_like.1.
 KM Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 236 AA; 30625 MW; 43CB86F44829CF3 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 296;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 AALAAVA 19
 |||||
 Db 216 AALAAVA 223

RESULT 30
 09CH13 PRELIMINARY; PRT; 306 AA.
 ID 09CH13;
 AC 09CH13;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Transcription regulator.
 GN YJFE OR L10928.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.,
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis sep. lactis IL1403."
 RL Genome Res. 11:731-753(2001).
 DR EMBL; AE006327; AAK05026.1; -
 DR InterPro: IPR004474; LytR_cpSA_psr.
 DR Pfam: PF03816; LytR_cpSA_psr.1.
 KM Complete proteome.
 SQ SEQUENCE 306 AA; 33657 MW; 2445A5CBA554CEBC CRC64;

Query Match 1.9%; Score 8; DB 16; Length 306;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 200 TLNAGDT 207
 |||||
 Db 202 TLNAGDT 209

RESULT 31
 09SNL7

ID 09SNL7 PRELIMINARY; PRT; 326 AA.
 AC 09SNL7;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Similar to magnesium-protoporphyrin IX methyltransferase (Putative
 DE magnesium-protoporphyrin IX methyltransferase).
 GN P0679C08.13.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.,
 RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 6, PAC
 RT clone:P0493C11."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.,
 RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 6, PAC
 RT clone:P0679C08."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002542; BAB19374.1; -
 DR Gramene; Q9SNL7; -
 DR InterPro: IPR000551; SAM bind.
 KM Methyltransferase; Transferase.
 SQ SEQUENCE 326 AA; 34926 MW; F8C91146F1772278 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 326;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AALAAAV 18
 |||||
 Db 65 AALAAAV 72

RESULT 32
 067350 PRELIMINARY; PRT; 348 AA.
 ID 067350;
 AC 067350;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Cation efflux system (CZCB-1like).
 GN CZCB2 OR AQ.1331.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPS;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M., Keller M., Aubay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus."
 RL Nature 392:353-358(1998).
 DR EMBL; AE000735; AAC07317.1; -
 DR InterPro: IPR005695; Membrane_fus2.
 DR InterPro: IPR001950; TIF_SUI1.
 DR TIGRFAMs: TIGR00999; 8a0102.1.
 DR PROSITE; PS01118; SUI1_1; 1.
 KM Complete proteome.
 SQ SEQUENCE 348 AA; 40570 MW; F4544EA6CA0635D CRC64;

Query Match 1.9%; Score 8; DB 16; Length 348;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 LKDGKRV 365
 |||||
 DB 325 LKDGKRV 332

RESULT 33

Q8GRS3 PRELIMINARY; PRT; 390 AA.

AC Q8GRS3; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OJ1477_F01.8 protein (P0047B07.31 protein).
 GN OJ1477_F01.8 OR P0047B07.31.
 OS Oryza sativa (japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC SpERMotophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhacroidae; Oryzaceae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
 clone: OJ1477_F01.8";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
 clone: P0047B07.31";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003833; BAC15474.1; -
 DR EMBL; AP005184; BAC16140.1; -
 SO SEQUENCE 390 AA; 44228 MW; BFD0C902BB861FAF CRC64;

Query Match 1.9%; Score 8; DB 10; Length 390;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVAL 20
 |||||
 DB 144 ALAAVAL 151

RESULT 34

Q9F241 PRELIMINARY; PRT; 394 AA.

AC Q9F241; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Putative membrane fusion protein.
 GN SMD.
 OS Xanthomonas maltophilia (Pseudomonas maltophilia) (Stenotrophomonas
 OS maltophilia).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Stenotrophomonas.
 OC NCBI_TaxID=40324;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN=D457R;
 RX MEDLINE=20493115; PubMed=11036026;
 RA Alonso A., Martinez J.L.;
 RT "Cloning and characterization of SmedF, a novel multidrug efflux pump
 from Stenotrophomonas maltophilia.";
 RL Antimicrob. Agents Chemother. 44:3079-3086 (2000).
 DR EMBL; AJ252200; CAC14594.1; -

DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SO SEQUENCE 394 AA; 40912 MW; D4234A037E47B53 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 394;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 RITAPISG 189
 |||||
 DB 171 RITAPISG 178

RESULT 35

Q8P7C8 PRELIMINARY; PRT; 399 AA.

AC Q8P7C8; 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Multidrug resistance protein.
 GN MEXA OR XCC2683.
 OS Xanthomonas campestris (pv. campestris).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xanthomonas.
 OC NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Forghieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Mejdans J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spholza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Serubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 host specificities.";
 RL Nature 417:459-463 (2002).
 DR EMBL; AE012380; AAM41955.1; -
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Complete proteome.
 SO SEQUENCE 399 AA; 59C1197DE6055D2D CRC64;

Query Match 1.9%; Score 8; DB 16; Length 399;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 AQATLAKA 129
 |||||
 DB 107 AQATLAKA 114

RESULT 36

Q8ZL01 PRELIMINARY; PRT; 408 AA.

AC Q8ZL01; 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Putative cytoplasmic protein.
 GN YIDR OR STM3811.
 OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SCS1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 Ryan E., Sun H., Florea L., Miller W., Stenking T., Nhan M.,
 Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 DR EMBL; AE008877; AL22670.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 408 AA; 45381 MW; 40454D052627FB08 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVTVHP 54
 |||||
 Db 64 VGVTVHP 71

RESULT 37
 Q822M0 PRELIMINARY; PRT; 408 AA.
 ID Q822M0;
 AC Q822M0;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Putative ATP/GTP-binding protein.
 GN STY3967.
 OS *Salmonella typhi*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgensen K.,
 Krogan A., Larsen T.S., Leach S., Moule S., O'Garra P., Parry C.,
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL; AL627280; CAD03183.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 408 AA; 45272 MW; 67EB07E7B9F7B43 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVTVHP 54
 |||||
 Db 64 VGVTVHP 71

RESULT 38
 Q8XBZ9 PRELIMINARY; PRT; 416 AA.
 ID Q8XBZ9;
 AC Q8XBZ9;

DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Orf, hypothetical protein.
 GN YIDR OR 25185 OR ECS4629.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 Kunara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AB005600; AAC58891.1; -
 DR EMBL; AP002566; BAB38052.1; -
 KW Complete proteome.
 SQ SEQUENCE 416 AA; 46314 MW; 37D16EB952A88669 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVTVHP 54
 |||||
 Db 76 VGVTVHP 83

RESULT 39
 Q8FBW4 PRELIMINARY; PRT; 416 AA.
 ID Q8FBW4;
 AC Q8FBW4;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein ydR.
 GN YIDR OR C4609.
 OS *Escherichia coli* O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22386234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 Raeko D., Buckles B.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AB016769; AA083044.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 416 AA; 46325 MW; 529AC5CD6C08FC30 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVVTVHP 54
 |||||
 |||||
 Db 76 VGVVTVHP 83

RESULT 40

O8G7V0 PRELIMINARY; PRT; 442 AA.
 ID O8G7V0;
 AC O8G7V0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Possible soluble binding protein of ABC transporter.
 GN BL0141.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=216816;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karimantzu M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
 RA Fridmore R.D., Arigoni F.;
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation
 RT to the human gastrointestinal tract."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL: AE014630; AAN24000.1; -.
 KW Complete proteome.
 SQ SEQUENCE 442 AA; 48635 MW; 1EBC99DACB45A9F CRC64;

Query Match 1.9%; Score 8; DB 16; Length 442;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LKDGDKVV 365
 |||||
 |||||
 Db 128 LKDGDKVV 135

Search completed: September 8, 2003, 14:10:47
 Job time : 70 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:03:22 ; Search time 17 Seconds

(without alignments)
1139.706 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412
Sequence: 1 MAFYAFKAMRAAALAAVAAL.....AAPQSGVOTASEAKTASEAE 412

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	21.6	412	1	MTRC_NEIGO
2	8	1.9	111	1	RL24_CHLMU
3	8	1.9	111	1	RL24_CHLTR
4	8	1.9	122	1	YM72_MYCTU
5	8	1.9	171	1	SFME_ECOLI
6	8	1.9	284	1	TRPA_AZOB
7	8	1.9	318	1	ARGB_BIFLO
8	8	1.9	416	1	CHRA_PSEAE
9	8	1.9	416	1	YIDR_ECOLI
10	8	1.9	434	1	UL43_HSV11
11	8	1.9	465	1	GA66_CHICK
12	8	1.9	478	1	AMP2_HUMAN
13	8	1.9	478	1	AMP2_MOUSE
14	8	1.9	478	1	AMP2_RAT
15	8	1.9	890	1	ADHE_ECOLI
16	8	1.7	91	1	MERP_SHIFL
17	8	1.7	98	1	YACB_RHISN
18	8	1.7	106	1	VH8B_BPT3
19	8	1.7	106	1	YNIU_THOSH
20	8	1.7	111	1	RL24_CHLPP
21	8	1.7	126	1	SYGB_NEIGO
22	8	1.7	126	1	TYRT_STRGA
23	8	1.7	134	1	RS12_HORVU
24	8	1.7	143	1	SYRB_HORVU
25	8	1.7	148	1	SYRB_RHISN
26	8	1.7	151	1	SYB2_RHIME
27	8	1.7	151	1	PSBU_CYACA
28	8	1.7	154	1	CYP2_ARATH
29	8	1.7	169	1	YSRO_STRCO
30	8	1.7	169	1	CYRH_CATRO
31	8	1.7	172	1	IF3_ANASP
32	8	1.7	177	1	IF3_THIET
33	8	1.7	180	1	IF3_THIET

34	7	1.7	182	1	KRUC_SHEEP	P26372	ovis aries
35	7	1.7	183	1	RBRF_MYCTU	P71614	mycobacteri
36	7	1.7	185	1	IF3_RICPR	Q92d19	rickettsia
37	7	1.7	186	1	IF3_RICCN	Q92hk6	rickettsia
38	7	1.7	187	1	Y418_VIBCH	Q9kuv7	vibrio chol
39	7	1.7	197	1	PAT_ALCPA	P31668	alcaligenes
40	7	1.7	200	1	COAE_BRUME	Q8ye21	brucella me
41	7	1.7	205	1	GSHC_BOVIN	Q09435	bos taurus
42	7	1.7	218	1	ESW2_DROME	Q97177	drosophila
43	7	1.7	226	1	OLE1_SOYBN	P29530	glycine max
44	7	1.7	232	1	SCOA_HELBO	Q921e3	helicobacte
45	7	1.7	251	1	DABP_CAMOE	Q9pit2	campylobact
46	7	1.7	252	1	GUNX_PRUP	P38534	prunus pers
47	7	1.7	267	1	IF2A_METRA	Q8xy86	methanopyru
48	7	1.7	270	1	HMPH_HUMAN	Q03014	homo sapien
49	7	1.7	271	1	HMPH_MOUSE	P43120	mus musculu
50	7	1.7	278	1	PROI_BACSU	P54552	baecillus su
51	7	1.7	281	1	KPRS_HALNI	Q9nh88	halobacteri
52	7	1.7	283	1	AROE_METTR	Q26144	archaeoglob
53	7	1.7	286	1	YL33_ARCRO	Q28147	archaeoglob
54	7	1.7	302	1	ARGB_STRCL	Q91e66	streplococc
55	7	1.7	302	1	WR40_ARATH	Q9esh7	arabidopsis
56	7	1.7	306	1	ARGB_STRCO	Q911a3	streplococc
57	7	1.7	312	1	HQ2_RABIT	P43242	oryctolagus
58	7	1.7	322	1	GDC_RAT	P16261	rattus norv
59	7	1.7	332	1	DNHJ_SYNY3	P50027	synecocyst
60	7	1.7	332	1	GDC_HUMAN	P16260	homo sapien
61	7	1.7	334	1	COBT_RHOCA	Q52679	rhodobacter
62	7	1.7	334	1	YL28_SCHRO	Q10343	schizosacch
63	7	1.7	344	1	MRAW_MYCTU	Q06212	mycobacteri
64	7	1.7	344	1	MURB_BUCAP	Q08463	buchnera ap
65	7	1.7	366	1	VUI_HSV6U	Q01349	human herpe
66	7	1.7	372	1	CYB_ELABA	Q92250	elaphe bair
67	7	1.7	372	1	FLGI_XANAC	Q8p127	xanthomonas
68	7	1.7	372	1	FLGI_XANCP	Q8p127	xanthomonas
69	7	1.7	372	1	LIG6_PHACH	Q8p127	xanthomonas
70	7	1.7	373	1	DDL_MYCTU	P50622	phanerocha
71	7	1.7	375	1	MURG_RHIL0	P50614	mycobacteri
72	7	1.7	376	1	TGT_RHIME	Q98d33	rhizobium l
73	7	1.7	380	1	MURG_BRUME	Q8y166	brucella me
74	7	1.7	386	1	CRFY_ERWHE	Q01331	erythrina her
75	7	1.7	400	1	TCR8_PASMU	P51564	pasteurella
76	7	1.7	405	1	TCR5_ECOLI	Q07282	escherichia
77	7	1.7	406	1	METK_AERPE	Q9ybk2	aeropyrum p
78	7	1.7	417	1	GLAI_PSEAE	Q9h7e5	pseudomonas
79	7	1.7	417	1	YAGI_YEAST	P33713	saccharomyc
80	7	1.7	418	1	GLA2_PSEAE	Q91138	pseudomonas
81	7	1.7	437	1	YHAM_ECOLI	Q42626	escherichia
82	7	1.7	441	1	VATH_ARATH	Q91x65	arabidopsis
83	7	1.7	456	1	KIM8_HUMAN	Q70135	homo sapien
84	7	1.7	458	1	A2AC_MOUSE	Q01337	mus musculu
85	7	1.7	458	1	A2AC_RAT	P22086	rattus norv
86	7	1.7	469	1	ARLY_BURUL	P56161	bukholderi
87	7	1.7	470	1	ESCA_DROME	P25932	drosophila
88	7	1.7	471	1	ARLY_PALSO	Q8xw17	raistonia s
89	7	1.7	471	1	KIM7_HUMAN	Q76014	homo sapien
90	7	1.7	477	1	BAIG_EUSBP	P33369	eubacterium
91	7	1.7	494	1	GUNI_PSEAE	P05552	persea amer
92	7	1.7	513	1	MATK_MAIZE	P48190	zea mays (m
93	7	1.7	515	1	ADCA_STRP3	Q08484	streplococc
94	7	1.7	515	1	ADCA_STRP8	Q8b1n2	streplococc
95	7	1.7	519	1	ADCA_STRPY	Q9a019	streplococc
96	7	1.7	519	1	DHAM_RAT	P11864	rattus norv
97	7	1.7	520	1	YM24_MYCTU	Q10509	mycobacteri
98	7	1.7	528	1	WR42_ARATH	Q92ec3	arabidopsis
99	7	1.7	538	1	WR31_ARATH	Q92ec3	arabidopsis
100	7	1.7	539	1	CH61_MYCTU	Q05973	mycobacteri
101	7	1.7	542	1	P69_MYCPN	P73369	mycoplasma
102	7	1.7	556	1	SRS4_CANAL	Q42816	candida alb
103	7	1.7	568	1	GGETI_PIG	P20735	sus scrofa
104	7	1.7	569	1	AAKH_HUMAN	Q9ugj0	homo sapien
105	7	1.7	590	1	HMAA_DROME	P25555	drosophila
106	7	1.7	597	1	CHIX_STR01	Q05638	streplococc

107	7	1.7	608	1	ATOS_ECOLI	Q06067	escherichia	180	6	1.5	109	1	SAS_PIG	Q29257	sus scrofa
108	7	1.7	621	1	DCTB_RHIME	P13633	rhizobium m	181	6	1.5	109	1	THYA_BOVIN	P01252	bos taurus
109	7	1.7	631	1	DMK_MOUSE	P5465	mus musculus	182	6	1.5	109	1	Y268_CAUCR	Q94658	caulobacter
110	7	1.7	631	1	ETPI_SCHPO	Q10361	schizosacch	183	6	1.5	110	1	PEPA_BOVIN	P00792	bos taurus
111	7	1.7	645	1	YI91_MYCTU	Q10384	mycobacteri	184	6	1.5	111	1	RU24_MYCPN	O50307	mycoplasma
112	7	1.7	659	1	MUTL_HAINT	Q9hame	halobacteri	185	6	1.5	113	1	HYPA_AZOCH	Q43948	azotobacter
113	7	1.7	681	1	DP3X_MYCPN	P75177	mycoplasma	186	6	1.5	113	1	HYPA_AZOVI	P31879	azotobacter
114	7	1.7	682	1	TDRS_HUMAN	Q8atc2	homo sapien	187	6	1.5	113	1	NUTI_WHEAT	P24296	triticum ae
115	7	1.7	687	1	SYGB_NEIMA	Q9j465	neisseria m	188	6	1.5	113	1	YEBY_ECOLI	P76272	escherichia
116	7	1.7	687	1	SYGB_NEIMB	Q9j467	neisseria m	189	6	1.5	119	1	AMCY_METEX	P04172	methylobact
117	7	1.7	688	1	SYGB_ECOL6	Q8xnt7	escherichia	190	6	1.5	119	1	PSPC_ECOLI	P23855	escherichia
118	7	1.7	688	1	SYGB_ECOLI	P09561	escherichia	191	6	1.5	120	1	CU19_ARADI	P80515	araneus dia
119	7	1.7	689	1	SYGB_YERPE	Q8z5w7	yersinia pe	192	6	1.5	121	1	VOR4_PIAWV	Q07520	plantago as
120	7	1.7	712	1	SMIA_CAEBL	P39396	caenorhabdi	193	6	1.5	122	1	RL14_CHLMU	Q9j474	chlamydia p
121	7	1.7	721	1	YJYI_ECOLI	P39396	caenorhabdi	194	6	1.5	122	1	RL14_CHLNP	Q9j477	chlamydia p
122	7	1.7	722	1	BPL1_MOUSE	Q920h2	m biotin--p	195	6	1.5	122	1	RL14_CHLNR	P28533	chlamydia t
123	7	1.7	774	1	RNFC_PSEAE	Q9hnb8	pseudomonas	196	6	1.5	123	1	RS12_BOPRU	O51348	borrelia bu
124	7	1.7	878	1	CAEP_PSEAE	Q9hxb3	pseudomonas	197	6	1.5	124	1	RS12_TREPA	Q8j271	treponema p
125	7	1.7	893	1	MVP_HUMAN	Q14764	homo sapien	198	6	1.5	125	1	THIH_PICWA	O63049	picea maria
126	7	1.7	904	1	SECA_RHOCA	P52966	rhodobacter	199	6	1.5	126	1	CLP2_DROMI	P91629	dtrosophila
127	7	1.7	904	1	HELY_MYCTU	Q10701	mycobacteri	200	6	1.5	126	1	FMKH_MICOL	Q08324	micromonsp
128	7	1.7	908	1	DM3A_MOUSE	Q88508	mus musculus	201	6	1.5	126	1	LY6E_CHICK	Q90986	gallus galli
129	7	1.7	909	1	DM3A_HUMAN	Q9y6k1	homo sapien	202	6	1.5	127	1	CU26_ARADI	P80517	araneus dia
130	7	1.7	917	1	IL6B_MOUSE	Q00560	mus musculus	203	6	1.5	127	1	VC41_HAEIN	P44235	haemophilus
131	7	1.7	918	1	IL6B_RAT	P40190	rattus norv	204	6	1.5	128	1	C562_ECOLI	P00192	escherichia
132	7	1.7	966	1	VIL3_ARATH	O81645	arabidopsis	205	6	1.5	128	1	C562_SALTY	O8x9q1	salmonella
133	7	1.7	1053	1	SLPM_BACBR	P06546	bacillus br	206	6	1.5	128	1	NIOA_MYCTU	P95181	mycobacteri
134	7	1.7	1062	1	CARB_BACHD	Q9kxv9	bacillus ha	207	6	1.5	134	1	Y353_PYRHO	O56791	pyrococcus
135	7	1.7	1144	1	CYV3_HUMAN	O60266	homo sapien	208	6	1.5	135	1	DEWA_EMENT	P52750	emeritella
136	7	1.7	1144	1	CYV3_RAT	P21932	rattus norv	209	6	1.5	138	1	PAR7_DICDI	P11022	dictyosteli
137	7	1.7	1145	1	CYV3_MOUSE	Q8vnh7	mus musculus	210	6	1.5	138	1	YCX1_CHLNU	O20118	chlorella v
138	7	1.7	1186	1	CEAA_BACTS	Q45710	bacillus th	211	6	1.5	139	1	Y965_MYCTU	P71545	mycobacteri
139	7	1.7	1212	1	SP22_HUMAN	P55011	homo sapien	212	6	1.5	141	1	RS12_METHH	O27129	methanobact
140	7	1.7	1302	1	GPSS_DROME	O61366	dtrosophila	213	6	1.5	143	1	MICR_RHIME	P55323	rhizobium m
141	7	1.7	1466	1	CA13_HUMAN	P02461	homo sapien	214	6	1.5	144	1	VC41_BPMV	P78680	bacterioph
142	7	1.7	2300	1	CYAA_NEUCR	Q01631	neutropora	215	6	1.5	144	1	Y911_METUA	Q58398	methanococ
143	7	1.7	3709	1	IRA2_YEAST	P19158	saccharomyc	216	6	1.5	146	1	HBH_ALUMI	P02130	alligator m
144	6	1.5	40	1	ANPB_MYOAE	P56571	rattus norv	217	6	1.5	146	1	HBH_CRONI	P03129	crocodilus m
145	6	1.5	40	1	ANPB_MYOAE	P20617	myxocephal	218	6	1.5	146	1	SP22_BACSH	O32724	bacillus sp
146	6	1.5	45	1	ANPB_MYOOSC	P04568	myxocephal	219	6	1.5	147	1	KL1B_STRLI	P22403	streptomyce
147	6	1.5	60	1	PER_METTL	P21305	methanococ	220	6	1.5	149	1	CYTM_HUMAN	Q15828	homo sapien
148	6	1.5	62	1	PERA_METUA	Q57652	methanococ	221	6	1.5	150	1	FABZ_AQUAE	O64668	aquifex aeo
149	6	1.5	63	1	NXS2_BUNFA	P14334	bungarus fa	222	6	1.5	150	1	PAL_PASMG	O51886	pasteurella
150	6	1.5	66	1	VHRI_VARY	P33797	variola vir	223	6	1.5	151	1	RS11_ANOGA	P52812	amphelies g
151	6	1.5	68	1	ATOX_HUMAN	O00244	homo sapien	224	6	1.5	151	1	SODC_SHEEP	P30670	ovis aries
152	6	1.5	68	1	ATOX_MOUSE	O08997	mus musculus	225	6	1.5	153	1	HOXT_AZOVI	P30781	azotobacter
153	6	1.5	68	1	ATOX_RAT	Q9wxc4	rattus norv	226	6	1.5	155	1	MAAZ_CHLFE	O8xgcs	chlorobium
154	6	1.5	72	1	LHAI_PROTE	P80588	rhodocyclu	227	6	1.5	155	1	RISB_CLOAB	Q91198	clostridium
155	6	1.5	72	1	RL29_TREPA	O83227	treponema p	228	6	1.5	155	1	SLYB_ECOLI	P55741	escherichia
156	6	1.5	78	1	RL28_ANASP	O8y6tc3	anabaena sp	229	6	1.5	155	1	SLYB_SALTY	O53549	salmonella
157	6	1.5	78	1	RL28_SYNY3	P72851	synecocyst	230	6	1.5	156	1	DTT_AGR75	O8u111	agrobacteri
158	6	1.5	79	1	PSPB_BOVIN	P15781	bos taurus	231	6	1.5	157	1	SSRP_BRAVA	Q9zsh74	bradyrhizob
159	6	1.5	80	1	COXM_HUMAN	P24411	homo sapien	232	6	1.5	159	1	SSRP_RHIL0	Q96539	rhizobium l
160	6	1.5	80	1	COXM_MOUSE	P56393	mus musculus	233	6	1.5	161	1	COAD_STRA3	Q866r1	streptococ
161	6	1.5	81	1	Y509_ECO57	P58092	escherichia	234	6	1.5	161	1	COAD_STRA5	Q861a6	streptococ
162	6	1.5	81	1	MOAD_ECOLI	P30748	escherichia	235	6	1.5	161	1	DTT_RHIL0	Q961a6	rhizobium l
163	6	1.5	84	1	MT22_LYCES	Q4315	lycopersico	236	6	1.5	161	1	MOAC_CLOAB	Q97h19	clostridium
164	6	1.5	85	1	RS17_MYCPN	O50309	mycoplasma	237	6	1.5	162	1	VC22_HAEIN	P44275	haemophilus
165	6	1.5	87	1	MOAC_EUBAC	Q93xd0	eubacterium	238	6	1.5	162	1	RS16_MYCTU	Q10795	mycobacteri
166	6	1.5	88	1	COXM_BOVIN	P13183	bos taurus	239	6	1.5	163	1	LEUD_PYRHO	O53393	pyrococcus
167	6	1.5	94	1	CH10_BRECH	O8ru01	brevibacill	240	6	1.5	164	1	TSTP_HSVSA	P18347	herpesvirus
168	6	1.5	95	1	MINE_DEIRA	Q9rub8	deinococcus	241	6	1.5	165	1	RSS_CHLFR	P28543	chlamydia t
169	6	1.5	96	1	RL28_AGR75	O8ubv8	agrobacteri	242	6	1.5	165	1	TPX_PSEAE	P57668	pseudomonas
170	6	1.5	98	1	CH10_BRUME	P25668	brucella me	243	6	1.5	167	1	YB20_AQUAE	O67200	aquifex aeo
171	6	1.5	100	1	YV74_CAUCR	Q45873	caulobacter	244	6	1.5	168	1	CYNS_ORISA	O67kx4	oryza sativ
172	6	1.5	101	1	PHS_RALSO	Q8xk38	raistonia s	245	6	1.5	171	1	FABZ_ANASP	O8yur4	anabaena sp
173	6	1.5	103	1	C552_NITEU	P95339	nitrosomona	246	6	1.5	172	1	COTX_BACSU	O08313	bacillus su
174	6	1.5	106	1	YK34_YEAST	P36092	saccharomyc	247	6	1.5	172	1	CYPL_ARATH	P34790	arabidopsis
175	6	1.5	107	1	CYV2_ORYSA	P20907	oryza sativ	248	6	1.5	173	1	NU6M_CALMA	P34196	calidris ma
176	6	1.5	107	1	PRV2_SALSA	O91883	salmo salar	249	6	1.5	173	1	NU6M_CHICK	P18941	gallus galli
177	6	1.5	107	1	PRVB_ESOLU	P02619	esox lucius	250	6	1.5	173	1	RUVC_VIBPA	O87qvc1	vibrio para
178	6	1.5	107	1	YU65_MYCTU	P95094	mycobacteri	251	6	1.5	174	1	SODC_CAUCR	P20379	caulobacter
179	6	1.5	108	1	RL24_MYCCE	P47408	mycoplasma	252	6	1.5	175	1	CD3E_CHICK	Q98910	gallus galli

253	6	1.5	175	1	VS10_ROTSL	P04512 simian 11 r	326	6	1.5	224	1	XP4_XENIA	Q00223 xenopus lae
254	6	1.5	176	1	HSIV_AQUAE	O67587 aquifex aeo	327	6	1.5	225	1	VSAS_RHME	O31185 rhizobium m
255	6	1.5	177	1	IL7_HUMAN	P13232 homo sapien	328	6	1.5	226	1	BIOD_RHME	O52587 mycobacteri
256	6	1.5	177	1	PRJQ_BACSU	P54173 bacillus su	329	6	1.5	226	1	PGMB_BACSU	O06995 bacillus su
257	6	1.5	179	1	ATPD_OCHNE	O46016 ochrosphaer	330	6	1.5	227	1	CIS3_YEAS	P47001 saccharomy
258	6	1.5	180	1	FMD3_NEIGO	Q00046 neisseria g	331	6	1.5	227	1	GPH_RHIL	O99m18 rhizobium i
259	6	1.5	180	1	YCBU_ECOLI	P75859 escherichia	332	6	1.5	227	1	GPRA_SPICT	P34030 spiropiasma
260	6	1.5	182	1	FMIA_ECOLI	P04128 escherichia	333	6	1.5	227	1	PSAF_CHLRE	P12356 chlamydomon
261	6	1.5	183	1	NER_HV1MN	P05852 human immun	334	6	1.5	228	1	VATE_MOUSE	P50518 mus musculu
262	6	1.5	183	1	YCF4_CHLYU	P56312 chlorella v	335	6	1.5	229	1	VG07_BP22	O01074 bacterioph
263	6	1.5	186	1	INB3_BOVIN	P01576 bos taurus	336	6	1.5	230	1	CUTI_FUSCO	P00590 fusarium so
264	6	1.5	186	1	INB3_BOVIN	P01577 bos taurus	337	6	1.5	230	1	CUTI_FUSCO	O99174 fusarium so
265	6	1.5	186	1	MAVE_PARVE	O56460 paracoccus	338	6	1.5	231	1	CUT3_FUSO	Q06u9 fusarium so
266	6	1.5	187	1	PAAD_CAMEB	Q09pfi campylobact	339	6	1.5	231	1	UCRI_NEUCR	P07056 neuropept
267	6	1.5	187	1	PYRE_THEMEA	Q09y96 thermotoga	340	6	1.5	233	1	G3PA_SINLA	P09672 sniaps alb
268	6	1.5	187	1	RUS2_THIFE	P24930 thiobacillu	341	6	1.5	233	1	MIP_LEGIO	P53605 legionella
269	6	1.5	188	1	COAT_CAVVU	P19128 cacao yello	342	6	1.5	234	1	VGGS_ECOLI	P53054 escherichia
270	6	1.5	189	1	NDXM_NEUCR	Q02854 neurospora	343	6	1.5	235	1	GSPN_PSEAE	Q51575 pseudomonas
271	6	1.5	189	1	VH02_VACCC	P20496 vaccinia vi	344	6	1.5	235	1	PYRF_AGRTS	P50638 agrobacteri
272	6	1.5	190	1	PYRE_THEMEA	P58858 thermoaer	345	6	1.5	236	1	GRA7_TOXGO	O00933 toxoplasma
273	6	1.5	190	1	RBS4_ECOLI	P13606 escherichia	346	6	1.5	237	1	PYRF_RHIL	O98dd5 rhizobium i
274	6	1.5	192	1	HIS7_AQUAE	O66455 aquifex aeo	347	6	1.5	238	1	PYRF_BRUME	O8yew9 bruceella me
275	6	1.5	192	1	YAGC_RHISN	P55459 rhizobium s	348	6	1.5	238	1	PYRF_BRUSU	O8fxw9 bruceella su
276	6	1.5	193	1	THM4_ARATH	Q09eue arabidopsis	349	6	1.5	239	1	GIDB_STRCO	O54571 streptomyc
277	6	1.5	193	1	VIN3_BPT4	P13302 bacterioph	350	6	1.5	240	1	RR40_YEAS	O08285 saccharomyc
278	6	1.5	194	1	VG37_HABIN	P44237 haemophilus	351	6	1.5	241	1	COAT_ABMWV	P21942 abutillon me
279	6	1.5	195	1	PAAD_DEIRA	Q9rr91 deinococcus	352	6	1.5	241	1	ENG8_BRUME	O8yda4 bruceella me
280	6	1.5	195	1	PGSA_PSEPL	P45419 pseudomonas	353	6	1.5	241	1	SPIR_SPIME	P21625 spiropiasma
281	6	1.5	195	1	TATB_XANCP	Q08jhn xanthomonas	354	6	1.5	241	1	UNG_AGRTS	O8ydw9 bruceella su
282	6	1.5	200	1	GTT1_DROSE	P30104 dirosophila	355	6	1.5	242	1	OMP4_NEITA	Q8u6m7 neisseria m
283	6	1.5	200	1	GTT1_DROMA	P30105 dirosophila	356	6	1.5	242	1	VMAT_INCUJ	P12446 influenza od
284	6	1.5	200	1	GTT1_DROSE	P30106 dirosophila	357	6	1.5	243	1	OMPA_SEROD	O53392 mycobacteri
285	6	1.5	200	1	GTT1_DROTE	P30107 dirosophila	358	6	1.5	244	1	YX42_MYCTU	O53392 mycobacteri
286	6	1.5	200	1	GTT1_DROTE	P30108 dirosophila	359	6	1.5	244	1	BDI_PASMU	O53392 mycobacteri
287	6	1.5	201	1	CLPP_CHLYU	P56317 chlorella v	360	6	1.5	245	1	CEIM_SALTY	Q05594 salmonella
288	6	1.5	201	1	CTFI_HUMAN	Q16619 homo sapien	361	6	1.5	245	1	HIS4_PSEBM	Q87ug3 pseudomonas
289	6	1.5	202	1	R13A_CAEBL	Q27389 caenorhadi	362	6	1.5	246	1	AGL5_ARATH	P29385 arabidopsis
290	6	1.5	202	1	YC55_MYCTU	Q11063 mycobacteri	363	6	1.5	246	1	PSA6_HUMAN	P34062 homo sapien
291	6	1.5	204	1	FIXJ_RHIME	P10958 rhizobium m	364	6	1.5	246	1	PSA6_MOUSE	Q9qum9 mus musculu
292	6	1.5	204	1	SN24_CARAV	P36977 carassius a	365	6	1.5	246	1	VC50_BACSU	P21968 bacillus su
293	6	1.5	204	1	TNE6_HUMAN	O95857 homo sapien	366	6	1.5	247	1	COBS_SALTY	Q05602 salmonella
294	6	1.5	205	1	RNH2_CAUCR	P52975 caulobacter	367	6	1.5	247	1	CYCX_DESDE	P81040 desulfovibr
295	6	1.5	206	1	KCY_RHOSH	Q9x4e0 rhodobacter	368	6	1.5	247	1	EFTS_SPIPL	P07477 spirulina p
296	6	1.5	207	1	VG66_BPMLS	O05280 mycobacteri	369	6	1.5	247	1	TRY1_HUMAN	P07477 spirulina p
297	6	1.5	209	1	GTT1_DROME	P20432 dirosophila	370	6	1.5	248	1	AGL1_ARATH	P29381 arabidopsis
298	6	1.5	210	1	ACPD_RHIL	Q98if8 rhizobium l	371	6	1.5	248	1	PHON_PROST	P26975 providencia
299	6	1.5	210	1	ALXH_TREPA	O85578 c putative	372	6	1.5	249	1	CISH_CHICK	Q9pw70 gallus galli
300	6	1.5	210	1	PSMB_METH	Q12970 methanobact	373	6	1.5	249	1	PHOC_MORNO	P28581 morganella
301	6	1.5	210	1	SAS_HUMAN	Q12999 homo sapien	374	6	1.5	250	1	ARGB_LISIN	Q92bb9 listeria in
302	6	1.5	212	1	BIOD_AGRTS	O8u8t9 agrobacteri	375	6	1.5	250	1	PSA4_ORYSA	O91e92 oryza sativ
303	6	1.5	212	1	GIDB_XYLFA	P37957 bacillus su	376	6	1.5	251	1	ATPD_PEA	O02758 plisum sativ
304	6	1.5	212	1	LIP_BACSU	O98c50 xyliella fas	377	6	1.5	251	1	CBX7_HUMAN	O95931 homo sapien
305	6	1.5	212	1	MSRA_VIBCH	Q9P30 vibrio chol	378	6	1.5	251	1	HIS6_SULSO	O33774 sulfobus
306	6	1.5	212	1	WFDI_RAT	O70280 rattus norv	379	6	1.5	252	1	YUG6_YEAS	P40374 saccharomyc
307	6	1.5	213	1	AROF_ERWHE	Q02285 erwina her	380	6	1.5	253	1	ADEFI_DROME	P05552 dirosophila
308	6	1.5	213	1	Y767_MYCTU	P71822 mycobacteri	381	6	1.5	254	1	PT22_YEAS	P10355 saccharomyc
309	6	1.5	215	1	YL97_MYCTU	Q10389 mycobacteri	382	6	1.5	255	1	MINC_RALSO	O8u920 ralsonia s
310	6	1.5	215	1	GTT2_DROME	Q9y998 dirosophila	383	6	1.5	255	1	SURE_RHOGE	Q9jp92 rhodocyclu
311	6	1.5	215	1	HMUO_CORDI	P71119 corynebacte	384	6	1.5	256	1	BRPA_STRHY	Q01108 streptomyc
312	6	1.5	215	1	Y093_TREPA	O83131 treponema p	385	6	1.5	256	1	DSBG_PSEAE	O91106 pseudomonas
313	6	1.5	216	1	GTT5_DROME	Q9y995 dirosophila	386	6	1.5	257	1	LPXA_FUSAN	O8ufu6 fusobacteri
314	6	1.5	216	1	UPP_SULSO	Q980G4 sulfolobus	387	6	1.5	258	1	ZM10_MOUSE	O54692 mus musculu
315	6	1.5	217	1	MTN_THEMEA	Q9y013 thermotoga	388	6	1.5	260	1	APAI_SPAU	O42175 sparus aur
316	6	1.5	219	1	R33B_BOVIN	P10948 bos taurus	389	6	1.5	260	1	PSAI_CHICK	O42265 gallus galli
317	6	1.5	219	1	R33B_RAT	Q63941 rattus norv	390	6	1.5	260	1	TH14_METAC	Q8tm19 methanosarc
318	6	1.5	220	1	ATPL_METUA	O57674 methanococc	391	6	1.5	262	1	FLGG_AGRTS	Q44338 agrobacteri
319	6	1.5	220	1	MINC_VIBVU	O8df54 vibrio vuln	392	6	1.5	262	1	UT11_CAEBL	O09462 caenorhadi
320	6	1.5	220	1	TPIS_CHLAU	P96744 chloroflexu	393	6	1.5	262	1	VN72_MYCTU	O05882 mycobacteri
321	6	1.5	220	1	TRBF_RHISN	P55403 rhizobium s	394	6	1.5	263	1	MURI_VIBVU	O8dd39 vibrio vuln
322	6	1.5	223	1	ATOA_HABIN	P44874 haemophilus	395	6	1.5	263	1	PSAI_HUMAN	P25786 homo sapien
323	6	1.5	224	1	PURQ_HALNI	Q9nmu2 halobacteri	396	6	1.5	263	1	PSAI_MOUSE	Q9r1p4 mus musculu
324	6	1.5	224	1	SAMP_MOUSE	P12246 mus musculu	397	6	1.5	263	1	PSAI_RAT	P18420 rattus norv
325	6	1.5	224	1	UREG_MYCTU	P50051 mycobacteri	398	6	1.5	264	1	FTSQ_STRCO	P45518 streptomyc

399	6	1.5	264	1	RPOD_ARCFU	O28002 archaeoglob	472	6	1.5	296	1	THIM_HUMAN	P25325 homo sapien
400	6	1.5	265	1	MURI_VIBCH	Q9KX17 vibrio chol	473	6	1.5	298	1	XERD_SALTY	P55889 salmoneilla
401	6	1.5	265	1	MYO2_LYCSC	P54927 lycopersico	474	6	1.5	299	1	NMS3_MYCTU	O10390 mycobacteri
402	6	1.5	265	1	RS2_LIRIAM	O43992 leishmania	475	6	1.5	299	1	PTB_BACSU	P54530 bacillus su
403	6	1.5	266	1	STR_KLEPN	P13082 klebsiella	476	6	1.5	300	1	V505_MYCLE	O49823 mycobacteri
404	6	1.5	267	1	LYLI_HUMAN	P12880 homo sapien	477	6	1.5	300	1	V759_CHLMU	O9P19 chlamydia m
405	6	1.5	268	1	ESI_HUMAN	P30042 homo sapien	478	6	1.5	302	1	RDGC_PASMU	P57810 pasteurella
406	6	1.5	268	1	ESLI_MYCGE	O49412 mycoplasma	479	6	1.5	302	1	YGBJ_ECOLI	O4888 escherichia
407	6	1.5	268	1	ESL2_MYCPN	P75311 mycoplasma	480	6	1.5	302	1	YHCT_BACSU	P54604 bacillus su
408	6	1.5	268	1	FKBA_ABRHY	O08437 aeromonas h	481	6	1.5	302	1	YH46_CLOAB	P58293 clostridium
409	6	1.5	268	1	MURI_TREPA	O83421 treponema p	482	6	1.5	304	1	DAPA_HAINT	O9P19 halobacteri
410	6	1.5	268	1	Y806_HAEIN	P44054 haemophilus	483	6	1.5	304	1	ICIR_PIG	O29000 sus scrofa
411	6	1.5	269	1	ZUPT_NEIMA	O91423 neisseria m	484	6	1.5	304	1	PHEA_AMEWE	O44104 amycolatops
412	6	1.5	269	1	ZUPT_NEIMB	O9K1H6 neisseria m	485	6	1.5	305	1	CDSA_METVA	O49433 m putative
413	6	1.5	270	1	PANB_OCEIH	O8E1F4 oceanobacil	486	6	1.5	305	1	HEM3_METVA	O84938 methanopyru
414	6	1.5	270	1	TRPC_RHOSH	O9Z6A7 rhodobacter	487	6	1.5	305	1	ILVE_SYNY3	P54691 synechocyst
415	6	1.5	271	1	KKAI_ECOLI	P00551 escherichia	488	6	1.5	305	1	PNP1_ANASP	O8YV19 anabaena sp
416	6	1.5	271	1	KKAI_SALTY	Q03447 salmoneilla	489	6	1.5	305	1	YEHZ_ECOLI	P33362 escherichia
417	6	1.5	271	1	THID_BACSU	P39610 bacillus su	490	6	1.5	306	1	DDG_ECOLI	P76522 escherichia
418	6	1.5	271	1	YNR3_AZOB	P45674 azospillulu	491	6	1.5	307	1	ERA_NEIMA	O91VD2 neisseria m
419	6	1.5	272	1	PANB_STPAM	O99365 staphylococ	492	6	1.5	307	1	ERA_NEIMB	P75931 escherichia
420	6	1.5	272	1	PANB_STPAM	O99365 staphylococ	493	6	1.5	307	1	MYIM_ECOLI	P06622 pseudomonas
421	6	1.5	272	1	YMAI_MYCBO	Q02277 mycobacteri	494	6	1.5	307	1	XYEI_PSEPU	P27887 pseudomonas
422	6	1.5	273	1	MURI_SHEON	O8E590 shewanella	495	6	1.5	307	1	XYLE_PSEAE	P27887 mycobacteri
423	6	1.5	273	1	PANB_RALSO	O8XW45 ralsstonia s	496	6	1.5	307	1	YMI6_MYCLE	O32960 mycobacteri
424	6	1.5	274	1	MADE_SPRCO	O91TMS streptomyce	497	6	1.5	307	1	YOJK_BACSU	P54548 bacillus su
425	6	1.5	276	1	PRXC_PSEPY	P25026 pseudomonas	498	6	1.5	309	1	PANB_MYCTU	O06280 mycobacteri
426	6	1.5	279	1	DAP2_ANASP	O8YV40 anabaena sp	499	6	1.5	309	1	TNMF_MOUSE	P41274 mus musculu
427	6	1.5	279	1	DAPF_SYNY3	P74667 synechocyst	500	6	1.5	309	1	V474_CHLTR	O84480 chlamydia t
428	6	1.5	279	1	PSAI_DROME	P12881 drosophila	501	6	1.5	310	1	PYRB_HAINT	O9H1H4 halobacteri
429	6	1.5	280	1	EROK_AARPE	O9YK6K aeropyrum p	502	6	1.5	310	1	WR18_ARATH	O9E5T4 arabidopsis
430	6	1.5	280	1	ETA_STPAU	P09331 staphylococ	503	6	1.5	310	1	V497_MYCTU	O11162 mycobacteri
431	6	1.5	281	1	AAC9_MYCCH	P29810 micromonospi	504	6	1.5	312	1	CYSK_SYNY3	P73410 synechocyst
432	6	1.5	281	1	PANB_MYCTU	O10505 mycobacteri	505	6	1.5	312	1	KPRS_XYLFA	O9PAP6 xyella fas
433	6	1.5	282	1	EFPS_ECOLI	P02997 escherichia	506	6	1.5	312	1	LIP_VIBCH	P15493 vibrio chol
434	6	1.5	282	1	PEG_MYCLE	O69470 mycobacteri	507	6	1.5	312	1	YCG9_METVA	O58686 methanococ
435	6	1.5	282	1	IFJ3_SCHPO	P78795 schizosacch	508	6	1.5	313	1	FIXB_ECOS7	O8A427 escherichia
436	6	1.5	282	1	Y8C2_PSEAE	O9HYV3 pseudomonas	509	6	1.5	313	1	FIXB_ECOLI	P53674 escherichia
437	6	1.5	283	1	DHPS_NEIMA	O9J170 neisseria m	510	6	1.5	313	1	FIXB_ECOLI	P13574 escherichia
438	6	1.5	283	1	DHPS_NEIMC	P57696 neisseria m	511	6	1.5	313	1	FIXB_SALTY	O8B405 salmoneilla
439	6	1.5	283	1	EFPS_SALTY	O8X860 salmoneilla	512	6	1.5	313	1	FIXB_SHIFL	P59675 shigella fl
440	6	1.5	283	1	MURI_SALTY	O8Z315 salmoneilla	513	6	1.5	313	1	FMT_THENA	O9WY28 thermotoga
441	6	1.5	283	1	MURI_SALTY	P40723 salmoneilla	514	6	1.5	314	1	MPR_BACSU	P33790 bacillus su
442	6	1.5	283	1	PANB_RHITO	O984D0 rhizobium 1	515	6	1.5	314	1	BLAC_STRAL	P14559 streptomyce
443	6	1.5	283	1	RPCI_BBP1	P13122 bacteriopho	516	6	1.5	314	1	IUNH_CRIPI	O27546 crithidia f
444	6	1.5	283	1	RPCI_BBP7	P13122 bacteriopho	517	6	1.5	314	1	MHPB_ECOLI	P54711 escherichia
445	6	1.5	284	1	AROK_HAINT	O9HBD0 halobacteri	518	6	1.5	316	1	FMT_CHLMU	O9P12 chlamydia m
446	6	1.5	284	1	TRBG_AGRTU	P54915 agrobacteri	519	6	1.5	316	1	T2BB_BACSU	P33562 bacillus su
447	6	1.5	284	1	VHRI_VACCC	P20632 vaccinia vi	520	6	1.5	317	1	ABE_PIG	P18650 sus scrofa
448	6	1.5	284	1	VHRI_VACCV	P04297 vaccinia vi	521	6	1.5	317	1	FMT_BACSU	P94463 bacillus su
449	6	1.5	285	1	MURI_ECOS7	O8X713 escherichia	522	6	1.5	318	1	KJPF_XANCP	P23354 xanthomonas
450	6	1.5	285	1	MURI_ECOL6	O8T689 escherichia	523	6	1.5	318	1	LDB_BACME	P00345 bacillus me
451	6	1.5	285	1	MURI_ECOLI	P22634 escherichia	524	6	1.5	318	1	RIPIA_TREPA	O83958 treponema p
452	6	1.5	285	1	PANB_CAUCR	O9AC88 caulobacter	525	6	1.5	319	1	K6PF_BACSH	O93114 bacillus sp
453	6	1.5	285	1	SYGB_HAEAE	O30836 haemophilus	526	6	1.5	320	1	RUD_PSEAE	P33640 pseudomonas
454	6	1.5	285	1	TI3B_HUMAN	O9Y754 homo sapien	527	6	1.5	320	1	YCFS_ECOLI	P73954 escherichia
455	6	1.5	286	1	CHM1_BRARE	P58239 brachydanio	528	6	1.5	321	1	ANX5_CHICK	P17153 gallus galli
456	6	1.5	286	1	DAPF_SYNEL	O8D4M2 synechococ	529	6	1.5	321	1	MOA1_HELP1	O94175 helicobacte
457	6	1.5	286	1	NIFC_CLOPA	P18795 clostridium	530	6	1.5	321	1	PGLB_BOVIN	P79119 bos taurus
458	6	1.5	286	1	YGBH_ECOLI	P11666 escherichia	531	6	1.5	321	1	PXAZ_RHIME	O9Z416 rhizobium m
459	6	1.5	286	1	YRAL_ECOLI	P45528 escherichia	532	6	1.5	321	1	PYRB_MYCLE	O9C455 mycobacteri
460	6	1.5	287	1	MURI_YERPE	O8ZAA2 yersinia pe	533	6	1.5	321	1	YMI1_BACSU	O03224 bacillus su
461	6	1.5	288	1	CRTE_RHOSH	P54976 rhodobacter	534	6	1.5	322	1	COAT_POPPV	O02106 poplar mosa
462	6	1.5	288	1	YB86_MYCTU	P71766 mycobacteri	535	6	1.5	322	1	ISPH_NEIMA	O61919 neisseria m
463	6	1.5	289	1	CRTE_RHOCA	P17060 rhodobacter	536	6	1.5	322	1	LIPA_RHIE1	O95941 rhizobium e
464	6	1.5	290	1	SUHB_MYCTU	O07703 mycobacteri	537	6	1.5	322	1	LIPA_RHIME	O9Z494 rhizobium m
465	6	1.5	290	1	Y873_METUA	O58283 methanococ	538	6	1.5	322	1	PELB_HUMAN	O96455 homo sapien
466	6	1.5	291	1	USF2_PRT	O63665 rattus norv	539	6	1.5	323	1	TKRA_ERWHE	P58000 erwinia her
467	6	1.5	292	1	EFPS_XYLFA	O9PAP6 xyella fas	540	6	1.5	324	1	ALGS_HUMAN	O9Y673 homo sapien
468	6	1.5	293	1	IF35_ARATH	O04202 arabidopsis	541	6	1.5	324	1	V640_METUA	O58057 methanococ
469	6	1.5	294	1	VG15_BPMLS	O05221 mycobacteri	542	6	1.5	326	1	YAMS_SCHPO	O10180 schizosacch
470	6	1.5	295	1	AMPN_PYRFU	P56218 pyrococcus	543	6	1.5	327	1	ACCA_SYNP7	O54766 synechococ
471	6	1.5	296	1	COPE_YEAST	P40509 saccharomyc	544	6	1.5	328	1	GLK_STAXY	O56198 staphylococ

545	6	1.5	328	1	GPDA_DEINA	Q9rr76	deinococcus	618	6	1.5	357	1	CAD7_PICAB	Q08350	picea abies
546	6	1.5	329	1	AN36_HELPY	P94851	heliobacte	619	6	1.5	357	1	COBT_BACHD	Q9rg31	baecillus ha
547	6	1.5	330	1	E13C_HORVU	Q02126	hordeum vul	620	6	1.5	357	1	OPSR_ASTFA	P22332	astyanax fa
548	6	1.5	330	1	FERG_ECOLI	P23877	escherichia	621	6	1.5	357	1	OPSR_CARAU	P13213	carassius a
549	6	1.5	330	1	GDG_BOVIN	Q01888	bos taurus	622	6	1.5	357	1	RPE_PASMU	Q9cnc8	pasturella
550	6	1.5	331	1	RRNG_HUMAN	Q9r644	homo sapien	623	6	1.5	359	1	ADD2_STRCO	Q9x722	streptomyce
551	6	1.5	332	1	G3P_PHYIN	P26988	phytophthor	624	6	1.5	359	1	COBT_CRYPA	P36562	escherichia
552	6	1.5	333	1	AOX2_SOYBN	Q41266	glycine max	625	6	1.5	359	1	GBB_CRYPA	O14435	cryptomonas
553	6	1.5	333	1	GLPX_HAEIN	P44811	haemophilus	626	6	1.5	359	1	OMPA_SERMA	P48445	serattia ma
554	6	1.5	334	1	ALG5_YEAST	P43550	saccharomyc	627	6	1.5	359	1	SLA9_MOUSE	Q08882	mus musculu
555	6	1.5	334	1	G3P_BACST	P00362	baecillus st	628	6	1.5	360	1	ALF_DROME	P07764	drosofila
556	6	1.5	334	1	G3P_CORGL	Q01651	corynebacte	629	6	1.5	360	1	CTG2_MAIZE	Q10717	zea mays (m
557	6	1.5	334	1	NAGZ_XANAC	Q08mu1	xanthomonas	630	6	1.5	360	1	FLGI_AQUAE	Q67608	aquilex aeo
558	6	1.5	335	1	G3P_BACCO	P15115	baecillus co	631	6	1.5	360	1	SYMM_MOUSE	Q9cyk1	mus musculu
559	6	1.5	335	1	TRXB_MYCTU	P52124	mycobacteri	632	6	1.5	360	-1	Y871_RICPR	Q9c995	ricketsia
560	6	1.5	336	1	RPOA_RHIL0	Q98n33	rhizobium l	633	6	1.5	361	1	IFP3_MOUSE	Q9ac44	mus musculu
561	6	1.5	337	1	HNR_ECOLI	P37055	escherichia	634	6	1.5	361	1	OPCT_HUMAN	Q16769	homo sapien
562	6	1.5	337	1	MRAM_CORGL	Q8nmn7	corynebacte	635	6	1.5	362	1	PROB_XANCP	Q088K2	xanthomonas
563	6	1.5	338	1	EFTS_BOVIN	P43896	bos taurus	636	6	1.5	363	1	PROB_XANAC	Q08K34	xanthomonas
564	6	1.5	338	1	MRAM_CORF	Q8fnt2	corynebacte	637	6	1.5	363	1	XYL2_PICST	P22144	picchia stip
565	6	1.5	338	1	TRPD_DEIRA	Q9ctj5	deinococcus	638	6	1.5	365	1	AROC_METAC	Q8ct87	methanosarc
566	6	1.5	338	1	VJHR_ECOLI	P39369	escherichia	639	6	1.5	365	1	ID12_METAC	Q8ct35	methanosarc
567	6	1.5	339	1	ERA_FICCN	Q92ja9	ricketsia	640	6	1.5	365	1	ID12_METMA	Q89w37	methanosarc
568	6	1.5	339	1	LIWC_CAUCR	Q96th4	caulobacter	641	6	1.5	367	1	CD5S_HUMAN	Q13319	homo sapien
569	6	1.5	339	1	PCB3_HUMAN	P57721	homo sapien	642	6	1.5	367	1	H181_RHIL0	Q987c8	rhizobium l
570	6	1.5	339	1	PCB3_MOUSE	P57722	mus musculu	643	6	1.5	368	1	H1SX_PYRAE	Q8xy17	pyrobaculum
571	6	1.5	340	1	GBB1_CAEEL	P17343	caenorhabdi	644	6	1.5	368	1	RH23_SCHPO	Q74803	schizosach
572	6	1.5	340	1	GBB3_CANFA	P79147	canis famli	645	6	1.5	369	1	CD5S_MOUSE	Q35926	mus musculu
573	6	1.5	340	1	GBB3_HUMAN	P15520	homo sapien	646	6	1.5	369	1	BR38_MEUCR	Q92249	neurospora
574	6	1.5	340	1	GBB3_MOUSE	Q61011	mus musculu	647	6	1.5	369	1	PSPD_BOVIN	P35246	bos taurus
575	6	1.5	340	1	GBB3_RAT	P52287	rattus norv	648	6	1.5	370	1	AROB_CAUCR	Q94434	caulobacter
576	6	1.5	340	1	HYPE_RHOCA	P80473	bos taurus	649	6	1.5	370	1	AROF_YEAST	P14843	saccharomyc
577	6	1.5	340	1	MYMS_BOVIN	P80473	bos taurus	650	6	1.5	370	1	RFL_MYCCA	P71496	mycoplasma
578	6	1.5	340	1	OTCA_LACPL	Q08322	lactobacill	651	6	1.5	371	1	CYB_COLCO	Q9m110	coluber con
579	6	1.5	340	1	YK4P_CAEEL	O17606	caenorhabdi	652	6	1.5	371	1	KLBI_ECOLI	P79558	dinodon sem
580	6	1.5	341	1	BUB3_YEAST	P26449	saccharomyc	653	6	1.5	372	1	CYB_DINSE	Q9g210	elaphe obso
581	6	1.5	342	1	G3P1_ANASP	P80506	anabaena sp	654	6	1.5	372	1	CYB_ELAVB	Q9g964	elaphe valp
582	6	1.5	342	1	G3P1_ANAVA	P34916	anabaena va	655	6	1.5	372	1	CYB_ELAVU	Q9k678	baecillus ha
583	6	1.5	342	1	HUPK_AZOVI	P31878	azotobacter	656	6	1.5	372	1	GPR_BACHD	Q95122	bos taurus
584	6	1.5	342	1	XYNA_CALSA	P23556	caldocellum	657	6	1.5	373	1	CD14_BOVIN	Q981	droseophila
585	6	1.5	343	1	HMAA_TRICA	Q07961	trihobium c	658	6	1.5	373	1	OPSI_DROME	P06002	thermoprote
586	6	1.5	343	1	NODI_RHISN	P55476	rhizobium s	659	6	1.5	373	1	Y38K_THERE	P05715	thermoprote
587	6	1.5	344	1	YO60_SYNY3	Q05153	synecocyst	660	6	1.5	374	1	CYSL_CAPAN	P13130	capasicum an
588	6	1.5	345	1	ESTA_STRSC	P22266	streptomyce	661	6	1.5	374	1	MURG_RHIME	Q92229	rhizobium m
589	6	1.5	345	1	FXLI_HUMAN	Q12952	homo sapien	662	6	1.5	375	1	AMSC_ERWAM	Q46633	erwinia amy
590	6	1.5	345	1	TRPD_AERPE	Q9r8c2	aeropyrum p	663	6	1.5	375	1	CD14_HUMAN	P08571	homo sapien
591	6	1.5	346	1	MURB_NEIMA	Q9y128	neisseria m	664	6	1.5	375	1	PWT1_TOBAC	Q42963	nicotiana t
592	6	1.5	346	1	MURB_NEIMB	Q9K016	neisseria m	665	6	1.5	375	1	TIAR_HUMAN	Q01085	homo sapien
593	6	1.5	346	1	USF2_HUMAN	Q15853	homo sapien	666	6	1.5	376	1	NIRK_RHIME	P316781	human papil
594	6	1.5	346	1	USF2_MOUSE	Q64705	mus musculu	667	6	1.5	376	1	VE2_HPV10	P50771	human papil
595	6	1.5	347	1	PLSX_RHIME	Q92qct5	rhizobium m	668	6	1.5	376	1	FLIH_POPTR	Q04064	populus tri
596	6	1.5	348	1	HOKV_AZOVI	P40597	azotobacter	669	6	1.5	377	1	GBB_KRATH	P49177	arabidopsis
597	6	1.5	349	1	MEFB_MOUSE	Q05087	mus musculu	670	6	1.5	378	1	FENI_MOUSE	P39749	mus musculu
598	6	1.5	349	1	RUVB_CLOAR	Q97qcl	clostridium	671	6	1.5	378	1	H182_CAUCR	Q945b6	caulobacter
599	6	1.5	351	1	APBE_ECOLI	P63943	escherichia	672	6	1.5	378	1	PEM1_PHACH	Q02567	planetococcae
600	6	1.5	351	1	B3AR_CAVPO	Q64085	cavia porce	673	6	1.5	378	1	BIOF_HAEIN	P44422	haemophilus
601	6	1.5	351	1	TRPF_LACLA	Q94775	homo sapien	674	6	1.5	380	1	CYB_MYSAL	Q917n5	myctomys a
602	6	1.5	352	1	TRPF_LACLA	P54314	mus musculu	675	6	1.5	380	1	BIOF_HAEIN	Q917n5	myctomys a
603	6	1.5	353	1	BMPC_BORBU	Q94b29	yersinia pe	676	6	1.5	379	1	CYB_MTCMC	P39749	mus musculu
604	6	1.5	353	1	CARA_HALNI	Q94b29	halobacteri	677	6	1.5	379	1	CYB_RHIME	P49177	arabidopsis
605	6	1.5	353	1	GBB5_HUMAN	Q94b29	halobacteri	678	6	1.5	379	1	BIOF_HAEIN	P49177	arabidopsis
606	6	1.5	353	1	GBB5_MOUSE	P54314	mus musculu	679	6	1.5	380	1	BIOF_HAEIN	P49177	arabidopsis
607	6	1.5	353	1	PMT2_TOBAC	Q94b29	halobacteri	680	6	1.5	380	1	BIOF_HAEIN	P49177	arabidopsis
608	6	1.5	353	1	TMPC_BURPA	P29724	treponema p	681	6	1.5	380	1	FENI_HUMAN	P49177	arabidopsis
609	6	1.5	354	1	MURG_BUCAP	Q849c4	buchnera ap	682	6	1.5	380	1	FTZ2_METUA	Q85039	methanococ
610	6	1.5	354	1	PON2_CHICK	Q90952	gallus gall	683	6	1.5	380	1	GBB_MAIZE	P49178	zea mays (m
611	6	1.5	354	1	PON2_MELGA	Q91090	melanogris g	684	6	1.5	380	1	GBB_ORYSA	Q40687	oryza sativ
612	6	1.5	354	1	RNFD_PSEST	Q94vna	pseudomonas	685	6	1.5	381	1	C3LI_MOUSE	Q61362	mus musculu
613	6	1.5	355	1	CARX_LACPL	Q08317	lactobacill	686	6	1.5	381	1	PWT3_TOBAC	Q94b29	haemophilus
614	6	1.5	355	1	Y497_MYCLE	P54580	mycobacteri	687	6	1.5	382	1	TALI_STRCO	Q08808	streptomyce
615	6	1.5	356	1	AROF_ECOLI	P00888	escherichia	688	6	1.5	382	1	OVAL_COTUA	P19104	coturnix co
616	6	1.5	356	1	AROF_SALTY	P21307	salmonella	689	6	1.5	382	1	Y894_HAEIN	Q57500	haemophilus
617	6	1.5	357	1	CAD2_PICAB	Q82035	picea abies	690	6	1.5	384	1	DNVJ_RHOCA	Q52702	rhodobacter

691	1.5	384	1	GUN_XANAC	P58935 xanthomonas	764	6	1.5	411	1	PGK_DEIRA	Q9rup2 deinococcus
692	1.5	384	1	Y823 MYCLE	Q50049 mycobacteri	765	6	1.5	412	1	ALF_PETHY	Q22621 petunia hyb
693	1.5	385	1	ACRE_ECOLI	P24180 escherichia	766	6	1.5	412	1	DAD3_RHIL0	Q981x2 rhizobium i
694	1.5	385	1	OVALL_CHICK	P01012 gallus gall	767	6	1.5	413	1	FU1_TOBAC	Q40504 nicotiana t
695	1.5	385	1	YEHY_ECOLI	P33361 escherichia	768	6	1.5	413	1	IF4E_TOBAC	P41381 nicotiana t
696	1.5	385	1	YH1U_ECOLI	P37636 escherichia	769	6	1.5	413	1	SCGE_MOUSE	Q70258 mus musculu
697	1.5	385	1	Y874_CLOAB	P45360 clostridium	770	6	1.5	414	1	IDH1_COLMA	P41560 colwella m
698	1.5	386	1	CYSL_SOLJU	O81155 solanum tub	771	6	1.5	414	1	SVH_MYCPN	P75069 mycoplasma
699	1.5	386	1	PHEA_ECOLI	P07022 escherichia	772	6	1.5	415	1	HUT1_RHIME	O31196 rhizobium m
700	1.5	386	1	TITAI_HUMAN	P31483 homo sapien	773	6	1.5	415	1	VEGM_ECOLI	P76397 escherichia
701	1.5	386	1	TITAI_MOUSE	P52912 mus musculu	774	6	1.5	416	1	FL2_TOBAC	O40505 nicotiana t
702	1.5	387	1	QIN_AVIS3	P56260 avian sarco	775	6	1.5	416	1	IDH_ECOLI	P08200 escherichia
703	1.5	388	1	RSGK_HUMAN	O76081 homo sapien	776	6	1.5	416	1	TFXD_RHIL0	P42726 rhizobium i
704	1.5	389	1	GPDI_ZYGRO	Q9hgyl zygosacchar	777	6	1.5	417	1	CZCC_ALCEU	P31509 alcaligenes
705	1.5	389	1	PKNS_MYCPN	P75524 mycoplasma	778	6	1.5	417	1	KCH_ECOLI	P31069 escherichia
706	1.5	389	1	Y823 MYCTU	O53835 mycobacteri	779	6	1.5	419	1	PMT4_TOBAC	Q9seha4 nicotiana t
707	1.5	390	1	TGFI_HORSE	O19011 equus caball	780	6	1.5	419	1	VC58_MYCTU	O11060 mycobacteri
708	1.5	390	1	YARI_RHISN	P55642 rhizobium s	781	6	1.5	419	1	YIBP_ECOLI	P37690 escherichia
709	1.5	391	1	IDH_STRSL	O59985 streptococc	782	6	1.5	420	1	MECC_SALTY	O82389 salmonella
710	1.5	391	1	P53_RAT	P10361 rattus norv	783	6	1.5	420	1	WECC_SALTY	O91674 salmonella
711	1.5	391	1	YDCO_ECOLI	P76103 escherichia	784	6	1.5	421	1	RL4_RAT	P50878 rattus norv
712	1.5	392	1	DXR_FORGL	O8np10 corynebacte	785	6	1.5	421	1	TOLA_ECOLI	P19934 escherichia
713	1.5	392	1	TITR_MOUSE	P70318 mus musculu	786	6	1.5	421	1	STY_EUCAP	P32135 escherichia
714	1.5	392	1	Y862 MYCLE	Q49682 mycobacteri	787	6	1.5	422	1	STY_EUCAP	O8a14 buchera ap
715	1.5	393	1	GUNI_USTWA	P54424 ustilago ma	788	6	1.5	423	1	DCOR_TRYBB	P07805 typanosoma
716	1.5	393	1	IDH_STRMU	O59940 streptococc	789	6	1.5	423	1	IDH_BACSU	P3126 bacillus su
717	1.5	393	1	OM45 YEAST	P16547 saccharomyc	790	6	1.5	423	1	PUR6_CAEEL	Q10457 c probabile
718	1.5	393	1	Y100_SVNY3	P54984 synechocyst	791	6	1.5	423	1	SVH_MYCTU	O50641 mycobacteri
719	1.5	394	1	CHS3_BROFI	O23729 bromheadia	792	6	1.5	424	1	CRTC_ORYSA	O981y8 oryza sativ
720	1.5	394	1	CHS4_BROFI	O23730 bromheadia	793	6	1.5	424	1	PRRC_STAMM	Q9aur7 scaphylococ
721	1.5	394	1	CHS8_BROFI	O23731 bromheadia	794	6	1.5	425	1	KAS1_STORM	P43678 streptomyce
722	1.5	394	1	Y6B9_PSEAE	P42314 pseudomonas	795	6	1.5	426	1	ASB4_MOUSE	Q9wv71 mus musculu
723	1.5	394	1	YEL8 YEAST	P39943 saccharomyc	796	6	1.5	426	1	DHMH_PARVE	P23006 paracoccus
724	1.5	395	1	UNI_FEA	O48559 piumu sativ	797	6	1.5	429	1	FLIK_BACSU	P23451 bacillus su
725	1.5	396	1	FLU_ANTMA	P23915 antirrhinum	798	6	1.5	429	1	THD1_MYCTU	Q10766 mycobacteri
726	1.5	396	1	G3PA_ARATH	P25856 arabidopsis	799	6	1.5	430	1	AKOA_MYCLE	Q9cc13 mycobacteri
727	1.5	396	1	ODPA_ASCSU	P26667 ascaris suu	800	6	1.5	430	1	PYRC_LACEL	P77884 lactobacilli
728	1.5	396	1	TCR3_ECOLI	P02981 escherichia	801	6	1.5	430	1	THIC_PYRAE	O8zxc0 pyrobaculum
729	1.5	397	1	ACRA_ECOLI	P31223 escherichia	802	6	1.5	432	1	ENO_TREPA	P74934 treponema p
730	1.5	397	1	CG2B_CARAU	O92162 carassius a	803	6	1.5	433	1	ELT2_CAEEL	Q10655 caenorhabdi
731	1.5	398	1	METK_METAC	O8tu57 methanosarc	804	6	1.5	433	1	RUC1_DICDI	O15746 dictyosteli
732	1.5	398	1	METK_METMA	O8pus4 methanosarc	805	6	1.5	434	1	ENO_DESVA	O32513 desulfovibr
733	1.5	398	1	PGK_STRAS	P20708 azotobacter	806	6	1.5	434	1	YU18_MYCTU	P51070 mycobacteri
734	1.5	398	1	PGK_STRAS	O8elf0 streptococc	807	6	1.5	435	1	IAAH_ENTAG	O50173 enterobacte
735	1.5	398	1	PGK_STRAS	O8elf0 streptococc	808	6	1.5	435	1	NMTC_BACSU	O53024 bacillus su
736	1.5	398	1	RA23 YEAST	O8dkt0 streptococc	809	6	1.5	435	1	YU21_MYCTU	O53268 mycobacteri
737	1.5	398	1	SPS_MOUSE	P32628 saccharomyc	810	6	1.5	436	1	ABD1_YEAST	P37383 saccharomyc
738	1.5	399	1	ARGD_RHIME	O9jfx2 mus musculu	811	6	1.5	436	1	ABGA_ECOLI	P77357 escherichia
739	1.5	399	1	TCRI_ECOLI	O928a0 rhizobium m	812	6	1.5	437	1	ARLY_CLOAB	O97ke5 clostridium
740	1.5	400	1	ARGD MYCTU	P94990 mycobacteri	813	6	1.5	437	1	PKNA_MYCLE	P54743 mycobacteri
741	1.5	400	1	METK_THEAC	O9hm12 thermoplasm	814	6	1.5	438	1	ARLY_CLOAB	P59616 clostridium
742	1.5	400	1	METK_THEVO	O97ct6 thermoplasm	815	6	1.5	438	1	SGCE_HUMAN	O43556 homo sapien
743	1.5	401	1	DXR_VIBPA	O87me3 vibrio para	816	6	1.5	438	1	SYD_PYRPU	O84256 pyrococcus
744	1.5	401	1	G3PA_SPTOL	P19666 spinacia ol	817	6	1.5	439	1	VCAC_LAMBD	P03711 bacteriophag
745	1.5	401	1	GPDI_ZYGRO	O9bqy2 zygosacchar	818	6	1.5	440	1	GPDI_YEAST	P41911 saccharomyc
746	1.5	401	1	METK_PYRPU	O8rwm1 pyrococcus	819	6	1.5	440	1	MURD_VIBCH	Q9K995 vibrio chol
747	1.5	402	1	ARGD STRCO	O91a4 streptomyce	820	6	1.5	440	1	YHJE_ECOLI	P37643 escherichia
748	1.5	402	1	HUT1_VIBCH	O9kq1 vibrio chol	821	6	1.5	441	1	DCTA_RHIME	P20672 rhizobium m
749	1.5	403	1	SV46_BRARE	O96t26 brachydanio	822	6	1.5	441	1	GSA_PROFR	O06774 propionibac
750	1.5	404	1	ARGD MYCLE	O9cc12 mycobacteri	823	6	1.5	442	1	DTA2_RHIL0	Q286r8 rhizobium i
751	1.5	404	1	KIME_SCHPO	O09780 schizosacch	824	6	1.5	442	1	IFR2_HUMAN	O12894 homo sapien
752	1.5	404	1	METK_PYRHO	O59488 pyrococcus	825	6	1.5	443	1	PO32_HUMAN	P20265 homo sapien
753	1.5	404	1	REPA_AGRRH	P05682 agrobacteri	826	6	1.5	443	1	Y878_MYCTU	O10540 mycobacteri
754	1.5	405	1	G3PA_PEA	P12858 pisum sativ	827	6	1.5	444	1	DCTA_RHIL0	O01857 rhizobium i
755	1.5	406	1	ARGD_LEPIN	P24087 leptospira	828	6	1.5	444	1	SHU4_ECOLI	P09748 escherichia
756	1.5	406	1	FBW3_HUMAN	O9ubd7 homo sapien	829	6	1.5	444	1	YVGO_YEAST	P47039 saccharomyc
757	1.5	407	1	WCAI_ECOLI	P32057 escherichia	830	6	1.5	445	1	PO32_MOUSE	P31360 mus musculu
758	1.5	408	1	ARGD_XANAC	O8p331 xanthomonas	831	6	1.5	445	1	PO32_RAT	P56622 rattus norv
759	1.5	408	1	ARGD_XANCP	O8p3c4 xanthomonas	832	6	1.5	446	1	ATPD_MYCLE	P51009 mycobacteri
760	1.5	409	1	KAS2_STRGA	P16339 streptomyce	833	6	1.5	446	1	DCUB_ECOLI	P14406 escherichia
761	1.5	409	1	YM65_MYCTU	O50697 mycobacteri	834	6	1.5	446	1	ENO_PLAFA	Q27727 plasmodium
762	1.5	410	1	PHT4_PSEPU	O05184 pseudomonas	835	6	1.5	446	1	MURF_STRO	O31804 streptomyce
763	1.5	411	1	DHMH_METEX	O49124 methyllobact	836	6	1.5	446	1	Y084_NEIMA	O9jx32 neisseria m

837	6	1.5	446	1	Y1B3_NEIME	Q9K1g9_neisseria_m	910	6	1.5	488	1	MURE_BRUSU	Q8fzpo_brucella_su
838	6	1.5	447	1	2ABG_RABIT	P50410_o_serine/ch	911	6	1.5	489	1	WR47_ARATH	Q9z8j1_arabidopsis
839	6	1.5	447	1	2ABG_RABIT	P97888_r_serine/ch	912	6	1.5	490	1	CPCI_HUMAN	P33260_homo_sapien
840	6	1.5	448	1	DC7A_AGRT5	P58373_agrobacteri	913	6	1.5	490	1	HDVD_CLOAM	P55792_clostridium
841	6	1.5	448	1	SDHD_PSEAE	O9hyn9_pseudomonas	914	6	1.5	490	1	TA53_TREDE	P18164_treponema_d
842	6	1.5	448	1	YD30_MYCTU	Q10841_mycobacteri	915	6	1.5	492	1	NYLA_FLASK	P13397_flavobacter
843	6	1.5	450	1	AROA_MYCTU	P22487_mycobacteri	916	6	1.5	492	1	NYLA_PSEB8	P13398_pseudomonas
844	6	1.5	451	1	FXGB_CHICK	Q90964_gallus_gall	917	6	1.5	493	1	RBL_GALSU	P13395_galdiera_s
845	6	1.5	451	1	YOE2_STRAT	Q53684_streptomyces	918	6	1.5	494	1	GATA_MYCTU	O53258_mycobacteri
846	6	1.5	452	1	GASR_RABIT	P46627_oryctolagus	919	6	1.5	494	1	NFED_PLEBO	O00239_plectonema
847	6	1.5	454	1	MURF_SYNY3	P45450_synchocyst	920	6	1.5	494	1	S61A_PYRSA	P18379_pyrenomonas
848	6	1.5	455	1	ARLY_CAUCR	Q94683_callobacter	921	6	1.5	495	1	ACCD_MYCTU	Q10861_mycobacteri
849	6	1.5	455	1	VI2_HPV11	P04013_human_papil	922	6	1.5	495	1	NFEL_KLEOX	P56267_klebsiella
850	6	1.5	457	1	ARLY_HAEIN	P44314_haemophilus	923	6	1.5	495	1	NFEL_KLEBN	P06772_klebsiella
851	6	1.5	458	1	GATD_METKA	Q8v884_methanopyru	924	6	1.5	495	1	NUSA_HAEIN	P43915_haemophilus
852	6	1.5	460	1	ARCD_ECOLI	P77429_escherichia	925	6	1.5	496	1	BAFI_KLWMA	P33293_kluyveromyc
853	6	1.5	460	1	II6A_MOUSE	P22272_mus_musculu	926	6	1.5	497	1	DC7A_RHISN	P13601_rhizobium_s
854	6	1.5	460	1	MURD_ENTHR	O07669_enterococcu	927	6	1.5	497	1	GATA_MYCLE	O33105_mycobacteri
855	6	1.5	460	1	SYC_PSEAE	Q942u7_pseudomonas	928	6	1.5	497	1	GLPK_LISTN	O93b6_listeria_in
856	6	1.5	461	1	DC7A_RHIGA	Q9x7K6_rhizobium_g	929	6	1.5	498	1	GLPK_LISTMO	O8y622_listeria_mo
857	6	1.5	461	1	DYR_SCHPO	P36591_schizosacch	930	6	1.5	498	1	YK04_MYCTU	Q10852_mycobacteri
858	6	1.5	462	1	II6A_RAT	P22273_rattus_novr	931	6	1.5	499	1	WETA_PENCH	O01870_penicillium
859	6	1.5	463	1	MURD_RHIME	Q52853_rhizobium_m	932	6	1.5	500	1	CILA_HAEIN	P44459_haemophilus
860	6	1.5	463	1	YB63_SYNY3	P73754_synchocyst	933	6	1.5	500	1	CPJ1_RABIT	P52786_oryctolagus
861	6	1.5	463	1	YB68_MYCTU	O53416_mycobacteri	934	6	1.5	501	1	KEPL_DROME	Q05652_drosophila
862	6	1.5	464	1	CAP_DICDI	P54654_dictyosteli	935	6	1.5	502	1	CPJ2_HUMAN	P51589_homo_sapien
863	6	1.5	465	1	FTSM_MYCLE	O50186_mycobacteri	936	6	1.5	503	1	YB97_MYCTU	O10818_mycobacteri
864	6	1.5	465	1	GID_DEIRA	Q9xru7_deinococcus	937	6	1.5	504	1	DC7P_ECOS7	O8x8y9_escherichia
865	6	1.5	466	1	MURD_CHLIE	Q8x9d2_chlorobium	938	6	1.5	504	1	DC7P_ECOLI	O8f113_escherichia
866	6	1.5	466	1	ARLI_RHILLO	Q98936_rhizobium_l	939	6	1.5	504	1	DC7P_ECOLI	O8f113_escherichia
867	6	1.5	466	1	ARLY_BRUSU	O8y1j7_brucella_me	940	6	1.5	504	1	DC7P_ESCPE	P13723_escherichia
868	6	1.5	466	1	ARLY_BRUSU	O8f4a4_brucella_su	941	6	1.5	504	1	DC7P_SALTY	O8z9b1_salmonella
869	6	1.5	466	1	ARLY_SYNP7	Q8g8t1_synchococc	942	6	1.5	504	1	DC7P_SALTY	P40733_salmonella
870	6	1.5	467	1	MURD_RHITO	Q981b1_rhizobium_l	943	6	1.5	504	1	DC7P_SHIBO	O59827_shigella_bo
871	6	1.5	467	1	MURD_BRUME	O8y168_brucella_me	944	6	1.5	504	1	ENGA_HAEIN	P44536_haemophilus
872	6	1.5	467	1	MURD_BRUSU	O8f182_brucella_su	945	6	1.5	505	1	CSOI_DROME	O27525_drosophila
873	6	1.5	469	1	FTSM_MYCTU	P71587_mycobacteri	946	6	1.5	505	1	LE12_METTH	O9y4t8_methanobact
874	6	1.5	471	1	ARLY_DEIRA	Q9rwj0_deinococcus	947	6	1.5	506	1	GATA_TREPA	O83983_treponema_p
875	6	1.5	471	1	GATA_THETH	Q91xc3_thermus_th	948	6	1.5	506	1	GUAA_DEIRA	O9z191_deinococcus
876	6	1.5	471	1	URO1_MAIZE	P16166_zea_mays_fm	949	6	1.5	507	1	GLK2_STRCO	O9fjm2_streptomyces
877	6	1.5	471	1	URO2_MAIZE	P16165_zea_mays_fm	950	6	1.5	508	1	GR83_ARATH	O22203_arabidopsis
878	6	1.5	471	1	UFO3_MAIZE	P16167_zea_mays_fm	951	6	1.5	508	1	GLPK_MYCOE	P47284_mycoplasma
879	6	1.5	472	1	SECD_SYNY3	Q05610_synchocyst	952	6	1.5	509	1	DLDD_MOUSE	O08749_mus_musculu
880	6	1.5	473	1	FLIT_AGRYS	O34171_agrobacteri	953	6	1.5	509	1	HMD1_ORYSA	P48019_oryza_sativ
881	6	1.5	474	1	MERA_STRLI	P30341_streptomyces	954	6	1.5	509	1	NADB_HALNI	O9hnm0_halobacteri
882	6	1.5	475	1	ORCA_SYNY3	P73720_synchocyst	955	6	1.5	510	1	CJ12_DROME	O9vnm6_drosophila
883	6	1.5	475	1	S611_CANFA	P38377_canis_famli	956	6	1.5	510	1	TAI_HUMAN	Q01101_homo_sapien
884	6	1.5	475	1	S611_HUMAN	P38378_homo_sapien	957	6	1.5	510	1	MURF_MYCTU	O06220_mycobacteri
885	6	1.5	475	1	S612_HUMAN	Q9z2x3_homo_sapien	958	6	1.5	511	1	INT_PSEAE	O9z186_pseudomonas
886	6	1.5	475	1	S612_MOUSE	O911x1_mus_musculu	959	6	1.5	512	1	HUTH_STRCO	O9z4w1_streptomyces
887	6	1.5	475	1	S61A_HALRO	O25147_haemocytthia	960	6	1.5	512	1	XYNA_CLOS	P13558_clostridium
888	6	1.5	477	1	DLDD_AZOVI	P18925_azotobacter	961	6	1.5	514	1	AT12_ARATH	O9zpj8_arabidopsis
889	6	1.5	477	1	FXGB_HUMAN	P55315_homo_sapien	962	6	1.5	514	1	CP77_ONCNY	CP77_ONCNY
890	6	1.5	477	1	HSS_RHOVI	O33233_rhodospseudo	963	6	1.5	514	1	TBLR_HUMAN	Q9zck7_homo_sapien
891	6	1.5	478	1	CLCA_YERPE	Q8abm0_yersinia_pe	964	6	1.5	515	1	CE04_DROME	O9z6c4_drosophila
892	6	1.5	479	1	VS12_TRYEB	P26327_trypanosoma	965	6	1.5	515	1	ROCA_BACSU	P36631_bacillus_su
893	6	1.5	480	1	DOC_RAT	P14173_rattus_novr	966	6	1.5	516	1	GUX1_NEMOR	P18676_neurospora
894	6	1.5	480	1	FXGB_RAT	Q00939_rattus_novr	967	6	1.5	516	1	RSP3_CHARE	P12759_chlamydomon
895	6	1.5	480	1	PTSB_STAYX	P51184_staphylococ	968	6	1.5	516	1	YI49_BRADA	Q45221_bradynithob
896	6	1.5	481	1	CB12_MOUSE	O8K3m5_mus_musculu	969	6	1.5	517	1	CP77_ORYZA	O9zsc3_rhizobium_m
897	6	1.5	481	1	FXGB_MOUSE	Q06087_mus_musculu	970	6	1.5	518	1	*HEMK_RICPR	O9zcb3_oryzias_lat
898	6	1.5	482	1	BPI_BOVIN	P17453_bov_taurus	971	6	1.5	518	1	NAEK_BRUME	O9zcb3_rickettsia
899	6	1.5	482	1	SYF_METTH	O26708_methanobact	972	6	1.5	518	1	SAP_CHICK	O8ybp2_brucella_me
900	6	1.5	483	1	NAHF_PSEBP	Q52460_pseudomonas	973	6	1.5	518	1	VI2_HPV47	O13035_gallus_gall
901	6	1.5	483	1	NDAD_ALCXX	P72349_alcaligenes	974	6	1.5	519	1	GSH1_YERPE	O24425_human_papil
902	6	1.5	484	1	YAVS_SCHPO	O10174_schizosacch	975	6	1.5	520	1	GUAA_RHIME	O9zsb3_rhizobium_m
903	6	1.5	484	1	YI64_VIBCH	Q9xk40_vibrio_choi	976	6	1.5	521	1	GUAA_XANCP	O8gk88_xanthomonas
904	6	1.5	485	1	YI64_HABIN	P71367_haemophilus	977	6	1.5	521	1	GUAA_XANCP	O8gk88_xanthomonas
905	6	1.5	486	1	CUG1_HUMAN	O9z879_homo_sapien	978	6	1.5	522	1	GUAA_XYLP4	O9z879_xylella_fas
906	6	1.5	486	1	CUG1_MOUSE	P28659_mus_musculu	979	6	1.5	524	1	CHID_BACCI	O9z879_bacillus_ci
907	6	1.5	486	1	SYC_STNEL	O8dhw6_synchococci	980	6	1.5	525	1	GUAA_MYCTU	O50729_mycobacteri
908	6	1.5	486	1	THRC_HELPY	O24924_helicobacte	981	6	1.5	525	1	GUAA_PSEAE	O9hnm6_pseudomonas
909	6	1.5	488	1	MIRE_BRUME	O8y171_brucella_me	982	6	1.5	525	1	NAEK_BRUSU	O8fwm5_brucella_su

```

983 6 1.5 525 1 VIAA_HUMAN Q9H598 homo sapien
984 6 1.5 525 1 VIAA_MACFA Q95K62 macaca fasc
985 6 1.5 525 1 VIAA_MOUSE Q35633 mus musculu
986 6 1.5 525 1 VIAA_RAT Q35458 rattus norv
987 6 1.5 526 1 EXTL_RHIME Q92877 rhizobium m
988 6 1.5 526 1 MURE_COREF Q8RT55 corynebacte
989 6 1.5 526 1 NADB_MYCLE Q9HY66 pseudomonas
990 6 1.5 527 1 GSH1_PSEAE Q9HY66 pseudomonas
991 6 1.5 527 1 NADB_MYCTU Q06595 mycobacteri
992 6 1.5 527 1 PUR1_MYCTU Q06626 mycobacteri
993 6 1.5 528 1 TCPO_MYCTU P39078 saccharomy
994 6 1.5 528 1 YC79_MYCTU Q11038 mycobacteri
995 6 1.5 529 1 SP15_STRGR P19471 streptomyce
996 6 1.5 530 1 VGLF_SVS P04849 simian viru
997 6 1.5 530 1 UDB2_RAT P08541 rattus norv
998 6 1.5 531 1 YP60_METTM P14933 methanobact
999 6 1.5 532 1 YMO0_YEAST Q04458 saccharomy
1000 6 1.5 533 1 APO2_STRCO Q04943 streptomyce

```

ALIGNMENTS

```

RESULT 1
MTRC_NEIGO STANDARD; PRT; 412 AA.
ID MTRC_NEIGO
AC P43505;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Membrane fusion protein mtrc precursor.
GN MTRC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=94254732; PubMed=8196548;
RA Pan W., Spratt B.G.;
RT "Regulation of the permeability of the gonococcal cell envelope by
the mtr system.";
RL Mol. Microbiol. 11:769-775(1994).
CC -1- FUNCTION: CELL MEMBRANE LIPOPROTEIN, INVOLVED IN CELL MEMBRANE
PERMEABILITY TO HYDROPHOBIC COMPOUNDS SUCH AS ANTIBIOTICS, DYES
AND DETERGENTS.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U14993; AAA80193.1; -
CC DR InterPro; IPR006143; H1yD.
CC DR Pfam; PF00529; H1yD; 1.
CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC KW Signal; Inner membrane; Lipoprotein.
CC FT CHAIN 1 24 POTENTIAL.
CC FT LIPID 25 412 MEMBRANE FUSION PROTEIN MTRC.
CC FT LIPID 25 25 N-ACYL DIGLYCERIDE (POTENTIAL).
CC SQ SEQUENCE 412 AA; 42773 MW; 97F9AFBCFAE321BA CRC64;

```

Query Match 21.6%; Score 89; DB 1; Length 412;
 Best Local Similarity 100.0%; Pred. No. 1e-77;
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 22 LSSCGKGDAAOGGPGAREAPAVGVYTHPTQVATLTPGLBSLRADVAQVGG 81
Db 22 LSSCGKGDAAOGGPGAREAPAVGVYTHPTQVATLTPGLBSLRADVAQVGG 81
Qy 82 IIOKRLFOEGSVVRGPLYQIDSTVEA 110
Db 82 IIOKRLFOEGSVVRGPLYQIDSTVEA 110

```

```

RESULT 2
RL24_CHLMTU STANDARD; PRT; 111 AA.
ID RL24_CHLMTU
AC Q9PUM5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR TC0804.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=83560;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gyll S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ullerbach T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AE002347; AAP39607.1; -
CC DR PIR; D81664; D81664.
CC DR PIR; E42645; E42645.
CC DR TIGR; TC0804; -
CC DR InterPro; IPR005824; KOW.
CC DR InterPro; IPR006466; KOW_sub.
CC DR InterPro; IPR003256; Ribosomal_L24.
CC DR InterPro; IPR005825; Ribosomal_L24_26.
CC DR Pfam; PF00467; KOW; 1.
CC DR ProDom; PD001677; Ribosomal_L24; 1.
CC DR SMART; SMO0739; KOW; 1.
CC DR TIGRFAMs; TIGR01079; rplX_bact; 1.
CC DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
CC KW Ribosomal protein; Complete proteome.
CC SQ SEQUENCE 111 AA; 12608 MW; 2F02863BA68C13F5 CRC64;

```

Query Match 1.9%; Score 8; DB 1; Length 111;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 DKVVEGI 369
 |||||
 Db 32 DKVVEGI 39

RESULT 3

```

RL24 CHLTR
ID RL24_CHLTR STANDARD; PRT; 111 AA.
AC P28537; 084523;
DT 01-DEC-1992 (Rel. 24, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR RL24 OR CTS17.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L2/434/Bu;
RX MEDLINE=92138612; PubMed=1735714;
RA Kaul R., Gray G.J., Koehncke N.R., Gu L.J.;
RT "Cloning and sequence analysis of the Chlamydia trachomatis spc
RT ribosomal protein gene cluster."
RL J. Bacteriol. 174:1205-1212(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99008089; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT. (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80325; AAA23173.1; -
DR EMBL; AE001323; AAC68118.1; -
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR003256; Ribosomal_L24.
DR InterPro; IPR005825; Ribosomal_L24_26.
DR Pfam; PF00467; KOW.1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR SMART; SM00739; KOW; 1.
DR TIGFAMS; TIGR01079; tpxX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein; Complete proteome.
FT VARIANT 85 R -> T (IN SEROVAR L2).
FT VARIANT 105 S -> L (IN SEROVAR L2).
FT SEQUENCE 111 AA; 12637 MW; 2F02805BC08C0165 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
DB 32 DKVVEGI 39

RESULT 4
ID YM72_MYCTU STANDARD; PRT; 122 AA.
AC Q50691;
DT 01-NOV-1997 (Rel. 35, Created)

```

```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein RV2272.
GN RV2272 OR MT2333 OR MTCY339.38C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gao S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Stulton J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayan L.A., Ernolava M.D., Salberg S.L.,
RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikus A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC -1- SIMILARITY: TO E.COLI Y1DH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77163; CAB00963.1; -
DR EMBL; AE007076; AAK46616.1; -
DR PIR; D70730; D70730.
DR TIGR; MT2333; -
DR InterPro; IPR003807; DUF202.
DR InterPro; IPR003807; DUF202.
DR Pfam; PF02656; DUF202; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
FT SEQUENCE 122 AA; 12994 MW; 23A53754264887A7 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LAAAVLV 21
DB 37 LAAAVLV 44

RESULT 5
ID SEMF_ECOLI STANDARD; PRT; 171 AA.
AC P38052; P75716; P77079;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fimbrial-like protein sfmf precursor.
 GN SFMF OR B0534.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 RA Federfisch N., Hyman R., Kaiman S., Komp C., Kurdi O., Lew H.,
 RA Lin D., Natch A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBD databases.
 RN [3]
 RP SEQUENCE OF 119-171 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90220507; PubMed=2183007;
 RA Muramatsu S., Mizuno T.;
 RT "Nucleotide sequence of the region encompassing the int gene of a
 RT cryptic prophage and the dna Y gene flanked by a curved DNA sequence
 RT of Escherichia coli K12.";
 RL Mol. Gen. Genet. 220:325-328(1990).
 RN [4]
 RP IDENTIFICATION
 RX MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC -1- SIMILARITY: STRONG, TO S.TYPHIMORIUM FIMF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000159; AAC73636.1; -;
 DR EMBL; U82598; AAB40732.1; ALT_INIT.
 DR EMBL; X51662; -; NOT_ANNOTATED_CDS.
 DR PIR; E64785; E64785.
 DR Ecogen; EG12388; sfmf.
 DR InterPro; IPR000259; Fimbrial.
 DR Pfam; PF00419; Fimbrial; 1.
 KM Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 171 FIMBRIAL-LIKE PROTEIN SFMF.
 FT DISULFID 40 80 POTENTIAL.
 SQ SEQUENCE 171 AA; 18222 MW; 59230981242AIDCA CRC64;

Query Match 1.9%; Score 8; DB 1; Length 171;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 LINAAGDTT 208
 |||||
 Db 61 LINAAGDTT 68

RESULT 6

TRPA_AZOBR STANDARD; PRT; 284 AA.
 ID TRPA_AZOBR
 AC O9LAG9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE tryptophan synthase alpha chain (EC 4.2.1.20).
 GN TRPA.
 OS Azospirillum brasilense.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 CC Rhodospirillaceae; Azospirillum.
 CC NCBI_TaxID=192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SP7;
 RX MEDLINE=20541289; PubMed=11092742;
 RA Dosselare F., Lambrecht M., Vanderleyden J.;
 RT "Isolation and sequence analysis of the trpBA gene cluster, encoding
 RT tryptophan synthase, from Azospirillum brasilense.";
 RL DNA Seq. 11:287-293(2000).
 CC -1- FUNCTION: The alpha subunit is responsible for the aldo cleavage
 CC of indoleglycerol phosphate to indole and glyceraldehyde 3-
 CC phosphate.
 CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
 CC = L-tryptophan + glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF139661; AAF61458.1; -;
 DR HSRP; P00929; 2MSY.
 DR HAMAP; MF_00131; -; 1.
 DR InterPro; IPR003009; FMN_enzyme.
 DR InterPro; IPR002028; trp_synthaseA.
 DR Pfam; PF00290; trp_synthA; 1.
 DR ProDom; PD001535; Trp_synthaseA; 1.
 DR TIGRFAMs; TIGR00262; trpA; 1.
 DR PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
 KW Tryptophan biosynthesis; Lyase.
 SQ SEQUENCE 284 AA; 29351 MW; EBB12C8727BFC57 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 284;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 SIAGITGA 377
 |||||
 Db 188 SIAGITGA 195

RESULT 7
 ARGB_BIFLO STANDARD; PRT; 318 AA.
 ID ARGB_BIFLO
 AC P59255;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acetylglutamate kinase (EC 2.7.2.8) (NAG kinase) (AGX) (N-acetyl-L-
 DE glutamate 5-phosphotransferase).
 GN ARGB OR BL1062.
 OS Bifidobacterium longum.
 CC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 CC Bifidobacteriaceae; Bifidobacterium.
 CC NCBI_TaxID=216816;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snell B., Vilanova D., Berger B.,
RA Pessi G., Zehren M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigon F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC -1- CATALYTIC ACTIVITY: ATP + N-acetyl-L-glutamate = ADP + N-acetyl-L-
CC glutamate 5-phosphate.
CC -1- PATHWAY: Arginine biosynthesis; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the acetylglutamate kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014729; AAN24870.1; -
DR HAMAP; MF 00082; -; 1.
DR Pfam; PF00696; aakkinase; 1.
DR TIGRFAMs; TIGR00761; argB; 1.
KW Arginine biosynthesis; Transferase; Kinase; Complete proteome.
FT ACT SITE 45 45 BY SIMILARITY.
FT ACT SITE 264 264 BY SIMILARITY.
FT SITE 100 100 SUBSTRATE BINDING (BY SIMILARITY).
FT SITE 102 102 SUBSTRATE BINDING (BY SIMILARITY).
FT SITE 203 203 SUBSTRATE BINDING (BY SIMILARITY).
SQ SEQUENCE 318 AA; 33818 MW; EC362CB1D32E1B7 CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 1; Length 318;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAAAAAV 18
DB 210 AAAAAAV 217

RESULT 8
CHRA_PSEAE STANDARD; PRT; 416 AA.
AC P14285;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Chromate transport protein.
GN CHRA.
OS Pseudomonas aeruginosa.
OG Plasmid pJM505.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90094230; PubMed=2152903;
RA Ceravantes C., Ohtake H., Chu L., Misra T.K., Silver S.;
RT "Cloning, nucleotide sequence, and expression of the chromate
RT resistance determinant of Pseudomonas aeruginosa plasmid pJM505."
RL J. Bacteriol. 172:287-291(1990).
CC -1- FUNCTION: THIS PROTEIN REDUCES CHROMATE ACCUMULATION AND IS
CC ESSENTIAL FOR CHROMATE RESISTANCE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- INDUCTION: By chromate.
CC -1- SIMILARITY: 29% IDENTITY TO ALCALIGENES EUTROPHUS CHROMATE
CC TRANSPORT PROTEIN CHRA.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29034; AAA88432.1; -
DR InterPro; IPR003370; Chromate_transp.
DR Pfam; PF02417; Chromate_transp; 2.
DR TIGRFAMs; TIGR00937; 2A51; 1.
KW Transports; Plasmid; Transmembrane; Chromate resistance.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 308 328 POTENTIAL.
FT TRANSMEM 341 361 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 395 415 POTENTIAL.
SQ SEQUENCE 416 AA; 44060 MW; 19F46F4D3B0D2479 CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 1; Length 416;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVAL 20
DB 213 ALAAVAL 220

RESULT 9
YIDR_ECOLI STANDARD; PRT; 416 AA.
ID YIDR_ECOLI
AC P31455; P76734;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yidR.
GN YIDR OR B3689.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication."
RL Genomics 16:551-561(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10328; AAA62041.1; -
DR EMBL; AE000446; AAC76712.1; -
DR PIR; B65171; B65171.
DR Ecocyc; Ecol1713; yidR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 416 AA; 46318 MW; 47E36360CC89A8B7 CRC64;

Query Match
1.9%; Score 8; DB 1; Length 416;

```


Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VGVWTVHP 54
DB 76 VGVWTVHP 83

RESULT 10

UL43_HSV11

ID UL43_HSV11 STANDARD; PRT; 434 AA.

AC P10227;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 01-DEC-1992 (Rel. 24, Last annotation update)

DE Membrane protein UL43.

GN UL43.

OS Herpes simplex virus (type 1 / strain 17).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Simplexvirus.

OX NCBI_TaxID=10299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88274327; PubMed=2839594;

RA McGeech D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,

RA McNab D., Perry L.J., Scott J.E., Taylor P.,

RT "The complete DNA sequence of the long unique region in the genome of

herpes simplex virus type 1."

RT J. Gen. Virol. 69:1531-1574(1988).

CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED OR TRANSMEMBRANE

CC PROTEIN (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL43,

EHV-1 17, AND VZV 15.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

RP SEQUENCE FROM N.A.

RC TISSUE=Embryonic brain;

RX MEDLINE=96373714; PubMed=8780005;

RA Bahn S., Harvey R.J., Darlison M.G., Wisden W.;

RT "Conservation of gamma-aminobutyric acid type A receptor alpha 6

subunit gene expression in cerebellar granule cells."

RT J. Neurochem. 66:1810-1818(1996).

CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE

VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE

GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE

CHANNEL.

CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)

RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

RP SEQUENCE FROM N.A.

RC TISSUE=Embryonic brain;

RX MEDLINE=96373714; PubMed=8780005;

RA Bahn S., Harvey R.J., Darlison M.G., Wisden W.;

RT "Conservation of gamma-aminobutyric acid type A receptor alpha 6

subunit gene expression in cerebellar granule cells."

RT J. Neurochem. 66:1810-1818(1996).

CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTRANSMITTER IN THE

VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE

GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE

CHANNEL.

CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)

RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
 OK NCBI_taxid=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95372350; PubMed=7644482;
 RA Artin S.W., Kendall R.L., Hall L., Weaver L.H., Stewart A.E.,
 RA Matthews B.W., Bradshaw R.A.;
 RT "Eukaryotic methionyl aminopeptidases: two classes of cobalt-dependent
 enzymes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95178556; PubMed=7873610;
 RA Li X., Chang Y.;
 RT "Molecular cloning of a human complementary DNA encoding an
 initiation factor 2-associated protein (p67).";
 RL Biochim. Biophys. Acta 1260:333-336(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Uscin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 109-478.
 RX MEDLINE=98030697; PubMed=9812898;
 RA Liu S., Widom J., Kemp C.W., Crews C.M., Clardy J.;
 RT "Structure of human methionine aminopeptidase-2 complexed with
 fumagillin.";
 RL Science 282:1324-1327(1998).
 CC -!- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 PROTEINS.
 CC -!- CATALYTIC ACTIVITY: Release of N-terminal amino acids,
 preferentially methionine, from peptides and arylamides.
 CC -!- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (By similarity).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U29607; AAA82930.1; -;
 DR EMBL; U13261; AAC63402.1; -;
 DR EMBL; BC013782; AAH13782.1; -;
 DR PIR; S52112; DPHUM2.
 DR PDB; 1B59; 14-JAN-00.
 DR PDB; 1B6A; 18-FEB-00.
 DR PDB; 1BN5; 31-JUL-99.
 DR PDB; 1BOA; 01-AUG-99.

DR MEROPS; M24.002; -;
 DR Genew; HGNC:16672; METAP2.
 DR MIM; 601870; -;
 DR GO; GO:0004239; P: methionyl aminopeptidase activity; TAS.
 DR GO; GO:0006464; P: protein modification; TAS.
 DR GO; GO:0006445; P: regulation of translation; TAS.
 DR InterPro; IPR002468; MAP_2.
 DR InterPro; IPR001714; Methamino_Pase.
 DR InterPro; IPR000994; Peptidase M24.
 DR Pfam; PF00557; Peptidase M24; 1.
 DR PRINTS; PR00599; MAPPEPTIDASE.
 DR TIGRfams; TIGR00501; met_pdae_II; 1.
 DR PROSITE; PS01202; MAP_2; 1.
 KW Hydroxylase; Aminopeptidase; Cobalt; 3D-structure.
 FT DOMAIN 36 46
 FT ACT_SITE 82 93
 FT DOMAIN 98 106
 FT ACT_SITE 231 231
 FT METAL 251 251
 FT METAL 262 262
 FT METAL 331 331
 FT METAL 364 364
 FT METAL 459 459
 FT METAL 120 123
 FT METAL 125 126
 FT STRAND 133 135
 FT HELIX 154 161
 FT TURN 162 162
 FT HELIX 163 186
 FT TURN 189 190
 FT STRAND 192 192
 FT HELIX 208 208
 FT TURN 209 209
 FT STRAND 211 211
 FT TURN 212 214
 FT STRAND 215 225
 FT TURN 226 227
 FT STRAND 230 230
 FT TURN 236 237
 FT STRAND 241 241
 FT TURN 244 245
 FT STRAND 248 256
 FT TURN 257 258
 FT STRAND 267 267
 FT HELIX 271 273
 FT TURN 274 290
 FT STRAND 293 294
 FT TURN 296 296
 FT STRAND 297 309
 FT HELIX 310 310
 FT TURN 312 315
 FT TURN 316 317
 FT STRAND 318 321
 FT STRAND 323 324
 FT TURN 326 327
 FT STRAND 329 334
 FT TURN 335 336
 FT STRAND 337 337
 FT STRAND 343 345
 FT STRAND 354 354
 FT TURN 357 358
 FT STRAND 360 368
 FT STRAND 375 376
 FT STRAND 382 385
 FT TURN 387 388
 FT HELIX 397 409
 FT TURN 410 412
 FT STRAND 415 416
 FT HELIX 417 422
 FT TURN 423 424
 FT HELIX 429 437
 FT TURN 438 439
 FT STRAND 441 444
 COBALT 2 (By similarity).
 COBALT 1 AND 2 (By similarity).
 COBALT 1 (By similarity).
 COBALT 1 (By similarity).
 COBALT 1 AND 2 (By similarity).

FT STRAND 447 448
 FT TURN 451 452
 FT STRAND 455 464
 FT STRAND 469 471
 FT TURN 472 473
 SQ SEQUENCE 478 AA, 52891 MW, 5788B4B83E48F9A CRC64;

Query Match 1.9%; Score 8; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 13
 AMP2_MOUSE STANDARD; PRT; 478 AA.
 ID AMP2_MOUSE 008663;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methionine aminopeptidase 2 (EC 3.4.11.18) (Metap 2) (Peptidase M 2)
 (Initiation factor 2 associated 67 kDa glycoprotein) (p67) (p67eIF2).
 CN METAP2 OR MNPEP OR P67EIF2.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN NCBI_TaxId=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RA Sekiguchi S., Suzuki E.;
 RL Submitted (Apr-1997) to the EMBL/Genbank/DDJ databases.
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 CC PROTEINS (BY SIMILARITY).
 CC -1- FUNCTION: PROTECTS EIF-2 ALPHA-SUBUNIT FROM INHIBITORY
 CC PHOSPHORYLATION BY EIF-2 KINASES. PLAYS A CRITICAL ROLE IN THE
 CC REGULATION OF PROTEIN SYNTHESIS. IT ALSO INTERACTS WITH THE
 CC EIF-2 GAMMA-SUBUNIT (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Release of N-terminal amino acids,
 CC preferentially methionine, from peptides and arylamides.
 CC -1- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, AB003144; BAA19789.1; -
 CC DR HSSP; P50579; 1B6A.
 CC DR MEROPS; M24.002; -
 CC MGD; MGI:1929701; Metap2.
 CC InterPro; IPR002468; MAP 2.
 CC InterPro; IPR001714; Mechanino_Phase.
 CC InterPro; IPR000994; Peptidase_M24.
 CC Pfam; PF00557; Peptidase_M24; 1.
 CC PRINTS; PR00599; MAPEPTIDASE.
 CC TIGRfams; TIGR00501; met_pdase_II; 1.
 CC DR PROSITE; PS01202; MAP 2; 1.
 CC DR Hydroxylase; Aminopeptidase; Cobalt.
 CC FT DOMAIN 36 46 ARG/LYS-RICH (BASIC).
 CC FT DOMAIN 82 93 ASP/GNU-RICH (ACIDIC).
 CC FT DOMAIN 98 106 POLY-LYS
 CC FT METAL 251 251 COBALT 2 (By similarity).
 CC FT METAL 262 262 COBALT 1 AND 2 (By similarity).
 CC FT METAL 331 331 COBALT 1 (By similarity).
 CC FT METAL 364 364 COBALT 1 (By similarity).

FT METAL 459 459 COBALT 1 AND 2 (By similarity).
 SQ SEQUENCE 478 AA; 52921 MW; BBB9A2AFc19952E8 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 14
 AMP2_RAT STANDARD; PRT; 478 AA.
 ID AMP2_RAT P38062;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methionine aminopeptidase 2 (EC 3.4.11.18) (Metap 2) (Peptidase M 2)
 (Initiation factor 2 associated 67 kDa glycoprotein) (p67) (p67eIF2).
 CN METAP2 OR MNPEP OR P67EIF2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxId=10116;
 RN NCBI_TaxId=10116;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Reuber H35; TISSUE=Liver;
 RC MEDLINE=93266517; PubMed=8496145;
 RA Wu S., Gupta S., Chatterjee N., Hileman R.E., Kinzy T.G.,
 RA Denlow N.D., Merrick W.C., Chakrabarti D., Osterman J.C., Gupta N.K.;
 RT "Cloning and characterization of complementary DNA encoding the
 RT eukaryotic initiation factor 2-associated 67-kDa protein (p67).";
 RL J. Biol. Chem. 268:10796-10801(1993).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RX MEDLINE=95372350; PubMed=7644482;
 RA Arfin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E.,
 RA Matthews B.W., Bradshaw R.A.;
 RT "Eukaryotic methionyl aminopeptidases: two classes of
 RT cobalt-dependent enzymes".
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 CC PROTEINS.
 CC -1- FUNCTION: PROTECTS EIF-2 ALPHA-SUBUNIT FROM INHIBITORY
 CC PHOSPHORYLATION BY EIF-2 KINASES. PLAYS A CRITICAL ROLE IN THE
 CC REGULATION OF PROTEIN SYNTHESIS. IT ALSO INTERACTS WITH THE
 CC EIF-2 GAMMA-SUBUNIT.
 CC -1- CATALYTIC ACTIVITY: Release of N-terminal amino acids,
 CC preferentially methionine, from peptides and arylamides.
 CC -1- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (By similarity).
 CC -1- PFM: CONTAINS 12 O-LINKED GLCNAC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, L10652; AAA41111.1; -
 CC DR PIR; A46702; A46702.
 CC DR HSSP; P50579; 1B6A.
 CC DR MEROPS; M24.002; -
 CC DR InterPro; IPR002468; MAP 2.
 CC DR InterPro; IPR001714; Mechanino_Phase.
 CC DR InterPro; IPR000994; Peptidase_M24.
 CC Pfam; PF00557; Peptidase_M24; 1.
 CC PRINTS; PR00599; MAPEPTIDASE.
 CC TIGRfams; TIGR00501; met_pdase_II; 1.

DR PROSITE: PS01202; MAP 2; 1.
 KW Hydrolyase; Aminopeptidase; Cobalt; Glycoprotein.
 FT DOMAIN 36 46 ARG/LYS-RICH (BASIC).
 FT DOMAIN 82 93 ASP/GLU-RICH (ACIDIC).
 FT METAL 98 106 POLY-LYS.
 FT METAL 251 251 COBALT 2 (By similarity).
 FT METAL 262 262 COBALT 1 AND 2 (By similarity).
 FT METAL 331 331 COBALT 1 (By similarity).
 FT METAL 364 364 COBALT 1 (By similarity).
 FT METAL 459 459 COBALT 1 AND 2 (By similarity).
 FT CONFLICT 464 478 LRPCKEYVSRGDDY -> CAQPVKLSAEHTIKT (IN REP. 1).
 SQ SEQUENCE 478 AA; 53052 MW; BEIC0B91E0CB3D74 CRC64;
 Query Match 1.9%; Score 8; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 DB 235 NAGDTTVL 242
 RESULT 15
 ID ADHE_ECOLI STANDARD; PRT; 890 AA.
 AC ADHE_ECOLI P17547;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aldohyde-alcohol dehydrogenase [includes: Alcohol dehydrogenase
 DE (EC 1.1.1.1) (ADH); Acetaldehyde dehydrogenase [acetylating]
 DE (EC 1.2.1.10) (ACDH); Pyruvate-formate-lyase deactivase (PFL
 DE deactivase)].
 GN ADHE OR ANA OR B1241 OR Z2016 OR ECS1741.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
 RA MEDLINE=90152365; PubMed=2695398;
 RA Goodlove P.E., Cunningham P.R., Parker J., Clark D.P.;
 RT "Cloning and sequence analysis of the fermentative alcohol-
 RT dehydrogenase-encoding gene of Escherichia coli.";
 RL Gene 85:209-214(1989).
 RN [2]
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-6, AND CHARACTERIZATION.
 RC STRAIN=K12;
 RC MEDLINE=91200315; PubMed=2015910;
 RA Kessler D., Leibrecht I., Knappe J.;
 RT "Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase
 RT activities of Escherichia coli reside on a polymeric protein particle
 RT encoded by adhE.";
 RL FEBS Lett. 281:59-63(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RC MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:123-147(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RC MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horichi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RC MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RC MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Onosubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE OF 848-890 FROM N.A.
 RC STRAIN=K12;
 RC MEDLINE=95291445; PubMed=7773397;
 RA Danchin A., Ktin E.;
 RT "Filling the gap between hns and adhE in Escherichia coli K12";
 RL Microbiology 141:959-960(1995).
 RN [8]
 RP SEQUENCE OF 1-20.
 RA MEDLINE=94359415; PubMed=7521508;
 RA Yamato M., Takahashi Y., Tomotake H., Ota F., Hirota K., Yamaguchi K.;
 RT "Monoclonal antibodies to spiroxin of Yersinia enterocolitica and
 RT analysis of the localization of spiroxin by use of chem.";
 RL Microbiol. Immunol. 38:177-182(1994).
 CC -1- FUNCTION: THIS ENZYME HAS THREE ACTIVITIES: ADH, ACDH, AND PFL-
 CC DEACTIVASE.
 CC -1- FUNCTION: PFL DEACTIVASE CATALYZES THE QUENCHING OF THE PYRUVATE-
 CC FORMATE-LYASE CATALYST IN AN IRON, NAD, AND COA DEPENDENT
 CC REACTION.
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
 CC NADH.
 CC -1- CATALYTIC ACTIVITY: Acetaldehyde + CoA + NAD(+) = acetyl-CoA +
 CC NADH.
 CC -1- COPACITOR: IRON ACTIVATED.
 CC -1- SUBUNIT: SEEMS TO FORM A ROD SHAPED POLYMER COMPOSED OF ABOUT 40
 CC IDENTICAL SUBUNITS.
 CC -1- INDUCTION: UNDER ANAEROBIC CONDITIONS IN THE ABSENCE OF NITRATE.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE IRON-
 CC CONTAINING ALCOHOL DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X59263; CA41195.1; -;
 CC EMBL; M33504; AAA23420.1; -;
 CC EMBL; AE000222; AAC74323.1; -;
 CC EMBL; D90759; BAA36121.1; -;
 CC EMBL; D90852; BAA16034.1; -;
 CC EMBL; AE005341; AAG56096.1; -;

OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCB1_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: NONE OBVIOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A5000067; AAB91632.1; -
DR PIR; T28629; T28629.
KW Hypothetical protein; Plasmid; Transmembrane.
FT TRANSMEM 2
FT TRANSMEM 70 90 POTENTIAL.
SQ SEQUENCE 98 AA; 10241 MW; B37332645061385 CRC64;

QY 11 AAAAIAA 17
Db 73 AAAAIAA 79

RESULT 18
VHSB_BPT3
ID VHSB_BPT3 STANDARD; PRT; 106 AA.
AC P20322;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Host specificity protein B.
GN 7.3.
OS Bacteriophage T3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCB1_TaxID=10759;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Luria;
RC MEDLINE=90133923; PubMed=2614843;
RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.";
RL J. Mol. Biol. 210:687-701(1989).
CC -1- FUNCTION: AFFECTS THE HOST RANGE (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17255; CAA35151.1; -
DR PIR; S07520; S07520.
RX STRAIN=CWL029;
SQ SEQUENCE 106 AA; 10842 MW; 5006890DF38A73A3 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 EAPAPV 47
Db 48 EAPAPV 54

RESULT 19
YNTU_RHOSH
ID YNTU_RHOSH STANDARD; PRT; 106 AA.
AC Q01195;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 10.8 kDa protein in nifu 5' region (ORF 1).
OS Rhodobacter sphaeroides (Rhodospirillum rubrum sphaeroides).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCB1_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283738; PubMed=1317839;
RA Meijer W.G., Tabita F.R.;
RT "Isolation and characterization of the nifusvW-zpon gene cluster from
RT Rhodobacter sphaeroides.";
RL J. Bacteriol. 174:3855-3866(1992).
CC -1- SIMILARITY: BELONGS TO THE HESB/YADR/YFHF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86823; AAA26135.1; -
DR PIR; A41880; A41880.
DR InterPro; IPR000361; HESB_YADR_YFHF.
DR Pfam; PF01521; HESB-1like; 1.
DR ProDom; PD002183; HESB_YADR_YFHF; 1.
DR TIGRFAMs; TIGR00049; TIGR00049; 1.
DR PROSITE; PS01152; HESB; 1.
KW Hypothetical protein.
SQ SEQUENCE 106 AA; 10780 MW; 5BAB3B3F7EADFC52 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 AAQAIRK 172
Db 7 AAQAIRK 13

RESULT 20
RL24_CHLNP
ID RL24_CHLNP STANDARD; PRT; 111 AA.
AC Q927R8; Q9J0G7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L24.
GN RPLX OR RL24 OR CPN0636 OR CP0111.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiaceae; Chlamydiales; Chlamydiales.
OX NCB1_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CWL029;
SQ MEDLINE=99206606; PubMed=10192388;

RA Kalkan S., Mitchell W., Marathe R., Iamuel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RT Nat. Genet. 21:385-389(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin W., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RT Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishi K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA.";
 RT Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: THIS PROTEIN IS FOUND IN THE RIBONUCLEOPROTEIN CORE AND
 CC IS INVOLVED IN THE EARLY ASSEMBLY OF THE 50S SUBUNIT. IT IS NOT
 CC INVOLVED IN THE FUNCTIONS OF THE MATURE 50S SUBUNIT (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001647; AAD18775.1; -;
 DR EMBL: AE002173; AAF37994.1; -;
 DR EMBL: AF002547; BAA98843.1; -;
 DR PIR: A86570; A86570.
 DR PIR: F72054; F72054.
 DR TIGR: CP0111; -;
 DR InterPro: IPR005824; KOW
 DR InterPro: IPR006646; KOW_sub.
 DR InterPro: IPR003256; Ribosomal_L24.
 DR InterPro: IPR005825; Ribosomal_L24_26.
 DR Pfam: PF00467; KOW; 1.
 DR ProDom: PD001677; Ribosomal_L24; 1.
 DR SMART: SM00739; KOW; 1.
 DR TIGRfam: TIGR01079; rplX_bact; 1.
 DR PROSITE: PS01108; RIBOSOMAL_L24; 1.
 DR Ribosomal protein, Complete proteome.
 SQ SEQUENCE 111 AA; 12425 MW; 7BF9C0021CD7358F CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 111;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 362 DKVVEG 368
 DB 32 DKVVEG 36
 RESULT 21
 RL20 CHIMU STANDARD; PRT; 123 AA.
 AC Q9PL84;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 50S ribosomal protein L20.
 GN RPLP OR TC0223.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Mopn / Nig;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin W., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RT Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50s ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE002289; AAF39095.1; -;
 DR PIR: E81727; E81727.
 DR TIGR: TC0223; -;
 DR HAMAP: MF_00382; -; 1.
 DR InterPro: IPR005813; L20.
 DR InterPro: IPR005812; L20_bact_org.
 DR Pfam: PF00453; Ribosomal_L20; 1.
 DR PRINTS: PR00062; RIBOSOMAL_L20.
 DR ProDom: PD002389; L20; 1.
 DR TIGRfam: TIGR01032; rplT_bact; 1.
 DR PROSITE: PS00937; RIBOSOMAL_L20; 1.
 DR Ribosomal protein, rRNA-binding, Complete proteome.
 SQ SEQUENCE 123 AA; 13911 MW; 8B5373CE03640024 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 174 AGINLR 180
 DB 86 AGINLR 92
 RESULT 22
 SYGB NEIGO STANDARD; PRT; 126 AA.
 ID SYGB NEIGO
 AC Q50945;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
 DE beta chain) (GlyRS) (Fragment).
 GN GLYS.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33084 / F62;
 RX MEDLINE=95053752; PubMed=7964493;
 RA Gotschlich E.C.;

```

RT "Genetic locus for the biosynthesis of the variable portion of
RT Neisseria gonorrhoeae lipooligosaccharide."
RL J. Exp. Med. 180:2181-2190 (1994).
CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA (Gly) = AMP + diphosphate
CC + glycyl-tRNA (Gly).
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14554; AAA68008.1; -.
DR HAMAP; MF_00255; -.
DR InterPro; IPR006194; tRNA_synth_Gly.
DR PROSITE; PS50861; AA tRNA_LIGASE II GUYAB; 1.
KW Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding.
FT NON TER 1
SQ SEQUENCE 126 AA; 13343 MW; C3913AD1A6591CF2 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAAA 17
Db 13 AAALAAA 19

RESULT 23
TYRT STRGA STANDARD; PRT; 134 AA.
AC P55047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tyrosinase co-factor (URF402).
GN MELC1.
OS Streptomyces glaucescens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1907;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / ETH 22794 / GLA.0;
RX MEDLINE=88040431; PubMed=3118334;
RA Huber M., Huettler R., Lerch K.;
RL "The promoter of the Streptomyces glaucescens mel operon.";
Nucleic Acids Res. 15:8106-8106 (1987).
RT -1- FUNCTION: THIS PROTEIN MAY FUNCTION TO DELIVER COPPER TO
TYROSINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00457; CA668512.1; -.
DR PIR; A26986; A26986.
DR InterPro; IPR006311; Tat.
DR TIGRfams; TIGR01409; Tat_signal_seq; 1.
KW Melanin biosynthesis; Copper.
SQ SEQUENCE 134 AA; 13593 MW; 1EB74145A7D1B9AE CRC64;

```

```

Query Match 1.7%; Score 7; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAAA 17
Db 13 AAALAAA 19

RESULT 24
RS12 HORVU
ID RS12 HORVU STANDARD; PRT; 143 AA.
AC Q9XHS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 40S ribosomal protein S12.
GN RPS12.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Himalaya; TISSUE=Leaf;
RA Robertson M., Asami T.;
RT "An unique S12 protein sequence from barley, Hordeum vulgare L. cv.
RT Himalaya, with sequence relatedness to animal S12 proteins of 40S
RT small ribosome subunit.";
Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE S12 FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF067732; AAD39838.1; -.
DR InterPro; IPR004038; Ribosomal L7A.
DR InterPro; IPR000530; Ribosomal S12e.
DR Pfam; PF01248; Ribosomal L7Ae; 1.
DR PRINTS; PR00972; RIBSOMAL_S12E; 1.
DR PROSITE; PS01189; RIBSOMAL_S12E; 1.
KW Ribosomal protein.
SQ SEQUENCE 143 AA; 15295 MW; 6A7263198992A880 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 EAAPAVV 47
Db 4 EAAPAVV 10

RESULT 25
SYRB RHISN
ID SYRB RHISN STANDARD; PRT; 148 AA.
AC P5535f;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-NOV-2000 (Rel. 39, Last annotation update)
DE Probable transcriptional regulator syrb.
GN SYRB OR Y4AN.
OS Rhizobium sp. (strain NGR234).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.,
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- FUNCTION: RESPONSIBLE FOR THE REPRESSION OF SYRM ACTIVITY
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SYRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000065; AAB91611.1; -
CC Nucleotide; Transcription regulation; Repressor; Plasmid.
CC SEQUENCE 148 AA; 16138 MW; F88BD9A19CD8979B CRC64;
SQ
Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 148;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 170 AIKSAGI 176
DB 78 AIKSAGI 84
RESULT 26
SYB2_RHIME
ID SYB2_RHIME STANDARD; PRT; 151 AA.
AC Q9Z3Q1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable transcriptional regulator syrb2.
GN SYRB2 OR RA0863 OR SWA1586.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CMX1-105;
RX MEDLINE=99413305; PubMed=10485295;
RA Sharypova L.A., Yurgel S.N., Keller W., Simarov B.V., Puhler A.,
RA Becker A.;
RT "The eff-482 locus of Sinorhizobium meliloti CMX1-105 that influences
RT symbiotic effectiveness consists of three genes encoding an
RT endoglucanase, a transcriptional regulator and an adenylate cyclase.";
RL Mol. Gen. Genet. 261:1032-1044(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Guzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- FUNCTION: SEEMS TO AFFECT THE TRANSCRIPTION OF CYA3. MAY BE
CC NEGATIVELY AUTOREGULATED.
CC -1- SIMILARITY: BELONGS TO THE SYRB FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITION 124 AND 137.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ225896; CAB38102.1; ALT_FRAME.
CC EMBL; AE007223; AAK65521.1; -
CC PIR; G95369; G95369.
CC InterPro; IPR002514; Transposase_8.
CC Pfam; PF01527; Transposase_8; 1.
CC Nucleotide; Transcription regulation; Repressor; Plasmid;
CC COMPLETE PROTEOME.
CC CONFLICT 17 N -> H (IN REF. 1).
CC SEQUENCE 151 AA; 17028 MW; 215627A0FAD4BBAB CRC64;
SQ
Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 151;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 170 AIKSAGI 176
DB 81 AIKSAGI 87
RESULT 27
SYB3_RHIME
ID SYB3_RHIME STANDARD; PRT; 151 AA.
AC P58346;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable transcriptional regulator syrb3.
GN SYRB3 OR RA0438 OR SWA0806.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Guzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- SIMILARITY: BELONGS TO THE SYRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE007234; AAK5096.1; -
CC PIR; F95316; F95316.
CC Nucleotide; Transcription regulation; Repressor; Plasmid;
CC COMPLETE PROTEOME.
CC SEQUENCE 151 AA; 16795 MW; 9CC59F4B86123B28 CRC64;
SQ
Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 151;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


OY 170 AIKSAGI 176
 DB 78 AIKSAGI 84

RESULT 28

PSBU CYACA STANDARD; PRT; 154 AA.
 ID PSBU CYACA 092055; 092780; 16-OCT-2001 (Rel. 40, Created)
 AC 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Photosystem II 12 kDa extrinsic protein, chloroplast precursor (PS II complex 12 kDa extrinsic protein) (PSII-U).
 GN PSBU.
 OS Cyanidium caldarium.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae; Cyanidium.
 OX NCBI_TaxID=2771;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99310600; PubMed=10381374;
 RA Ohta H., Okumura A., Okuyama S., Akiyama A., Iwai M., Yoshinara S., Shen J.-R., Kamo M., Enami I.,
 RT "Cloning, expression of the psbu gene, and functional studies of the recombinant 12-kDa protein of photosystem II from a red alga
 RT Cyanidium caldarium";
 RL Biochem. Biophys. Res. Commun. 260:245-250 (1999).
 RN [2]
 RP SEQUENCE OF 62-112.
 RX MEDLINE=96124997; PubMed=8534673;
 RA Enami I., Murayama H., Ohta H., Kamo M., Nakazato K., Shen J.-R.;
 RT "Isolation and characterization of a Photosystem II complex from the red alga Cyanidium caldarium: association of cytochrome c-550 and a 12 kDa protein with the complex";
 RL Biochim. Biophys. Acta 1232:208-216 (1995).
 CC -1- FUNCTION: Constitutively stabilizes and protects the oxygen-evolving complex of photosystem II against heat-induced
 CC inactivation (by similarity).
 CC -1- SUBCELLULAR LOCATION: Loosely bound to the inner surface of the chloroplast thylakoid membrane.
 CC -1- SIMILARITY: BELONGS TO THE PSBU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AB023805; BAA75398.1; AUT_INIT.
 DR TIGRFAAS; TIGR01409; TAT_signal_seq; 1.
 KW Photosynthesis; Photosystem II; Thylakoid; Membrane; Transit peptide; Chloroplast.
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT TRANSIT 2 ? THYLAKOID.
 FT CHAIN 62 154 PHOTOSYSTEM II 12 KDA EXTRINSIC PROTEIN.
 FT CONFLICT 80 80 N -> D (IN REF. 2).
 FT CONFLICT 90 90 F -> I (IN REF. 2).
 SQ SEQUENCE 154 AA; 16715 MW; EF28D01ABEC0977F CRC64;

Query Match 1.7%; Score 7; DB 1; Length 154;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ALAAVA 19
 DB 44 ALAAVA 50

RESULT 29

CYP2 ARATH STANDARD; PRT; 169 AA.
 ID CYP2 ARATH P35627;
 AC 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Notamase) (Cyclophilin) (Cyclosporin A-binding protein).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta; TISSUE=leaf;
 RX MEDLINE=92322989; PubMed=1623198;
 RA Bartling D., Heese A., Weller E.W.;
 RT "Nucleotide sequence of a cDNA encoding an Arabidopsis
 RT cyclophilin-like protein";
 RL Plant Mol. Biol. 19:529-530 (1992).
 CC -1- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides.
 CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline (omega=0).
 CC -1- ENZYME REGULATION: BINDS CYCLOSPORIN A (CSA). CSA MEDIATES SOME OF ITS EFFECTS VIA AN INHIBITORY ACTION ON PPIASE.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -1- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X63616; CAA45161.1; -.
 DR PIR, S22496; S22496.
 DR HSSP; P05092; 2CPL.
 DR InterPro; IPR002130; CSA_PPIase.
 DR Pfam; PF00160; pro_isomerase; 1.
 DR PROSITE; PRO0153; CSAPISMASE.
 DR PROSITE; PS00170; CSA_PPIASE_1; 1.
 DR PROSITE; PS50072; CSA_PPIASE_2; 1.
 CC Cyclosporin; Isomerase; Rotamase; Multigene family.
 SQ SEQUENCE 169 AA; 18161 MW; B6077FC139864931 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 169;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 GGOPAGR 40
 DB 11 GGOPAGR 17

RESULT 30

YSFO STROO STANDARD; PRT; 169 AA.
 ID YSFO STROO 092863; 28-FEB-2003 (Rel. 41, Created)
 AC 092863; 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Hypothetical protein SC06450.
 GN SC06450 OR SC985.17.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; OC Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kisser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieker T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Watzdorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RT Nature 417:141-147(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL035206; CAA22759.1; -
DR PIR; T35937; T35937.
DR HAMAP; MF_01205; -, 1.
DR InterPro; IPR002589; A1pp.
DR Pfam; PF01661; A1pp; 1.
DR SMART; SM00506; A1pp; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 169 AA; 17767 MW; 86965DB33C6E4740 CRC64;
SQ
Query Match 1.7%; Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 68 ESLRTAD 74
Db 103 ESLRTAD 109
RESULT 31
CYPH CATRO STANDARD; PRT; 172 AA.
AC Q39613;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase)
DE (Cyclophilin) (Cyclosporin A-binding protein).
GN PCKR1
OS Catharanthus roseus (Rosa periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Gentianales; Apocynaceae; Rauvolfioideae; Vincaceae;
OC Catharanthus.
OC NCBI_TaxID=4058;
OK NCBI_TaxID=4058;
RN [1]
RP SEQUENCE FROM N.A.
RA Claessens M., Maeroufi H., Andreu F., Chentleux J.-C., Rideau M.,
RA Hamdi S.;
RT "Isolation of a full-length cDNA encoding a cytosolic cyclophilin from
RT periwinkle (Catharanthus roseus).";
RT (in) Plant Gene Register PGR95-100.
CC -1- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides.
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -1- ENZYME REGULATION: BINDS CYCLOSPORIN A (CSA). CSA MEDIATES SOME
CC OF ITS EFFECTS VIA AN INHIBITORY ACTION ON PPIASE.

```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X85185; CAA59468.1; -
DR PIR; T10056; T10056.
DR HSSP; P05092; 1CWL.
DR InterPro; IPR002130; CSA PPIase.
DR Pfam; PF00160; pro isomerase; 1.
DR PRINTS; PR00153; CSA_PPIASE.
DR PROSITE; PS00170; CSA_PPIASE_1; 1.
DR PROSITE; PS50072; CSA_PPIASE_2; 1.
KW Cyclosporin; Isomerase; Rotamase; Multigene family.
SQ SEQUENCE 172 AA; 18285 MW; EA6EC51886A50A81 CRC64;
Query Match 1.7%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 34 GGQAPGR 40
Db 13 GGQAPGR 19
RESULT 32
IF3 ANASP STANDARD; PRT; 177 AA.
ID IF3 ANASP
AC Q8WTE3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Translation initiation factor IF-3.
GN INFC OR ALL4623.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Rep. 8:205-213(2001).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP003597; BAB76322.1; ALT_INIT.
DR HAMAP; MF_00080; -, 1.
DR InterPro; IPR001288; IF3.

```

DR Pfam; PF00707; IF3_C; 1.
 DR Pfam; PF05198; IF3_N; 1.
 DR ProDom; PD002880; IF3; 1.
 DR TIGRFAMs; TIGR00168; infc; 1.
 DR PROSITE; PS00938; IF3; 1.
 SK Initiation factor; Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 177 AA; 20750 MW; 9C1585FF62590EFP CRC64;

Query Match 1.7%; Score 7; DB 1; Length 177;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LKDGDKV 364
 |||||
 DB 118 LKDGDKV 124

RESULT 33

IF3_THETN STANDARD; PRT; 180 AA.
 AC Q8R9C2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Translation initiation factor IF-3.
 GN INFC OR TTE1693.
 OS Thermobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermobacteriales;
 OC Thermobacteriaceae; Thermobacter.
 OK NCBI_TaxID=119072;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 Chen Y., Xue Y., Xu Y., Lai X., Huang L.,
 Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome.";
 RL Genome Res. 12:689-700(2002).

CC -!- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
 equilibrium between 70S ribosomes and their 50S and 30S subunits in
 favor of the free subunits, thus enhancing the availability of 30S
 subunits on which protein synthesis initiation begins.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the IF-3 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AE013124; AAM24894.1; -.
 DR HAMAP; MF_00080; -; 1.
 DR InterPro; IPR001288; IF3.
 DR Pfam; PF00707; IF3_C; 1.
 DR Pfam; PF05198; IF3_N; 1.
 DR ProDom; PD002880; IF3; 1.
 DR TIGRFAMs; TIGR00168; infc; 1.
 DR PROSITE; PS00938; IF3; 1.
 KW Initiation factor; Protein biosynthesis; Complete proteome.
 SO SEQUENCE 180 AA; 21061 MW; 8398335388F8CEBB CRC64;

Query Match 1.7%; Score 7; DB 1; Length 180;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LKDGDKV 364
 |||||
 DB 120 LKDGDKV 126

RESULT 34

KRUC_SHEEP STANDARD; PRT; 182 AA.
 AC P26372;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Keratin, ultra high-sulfur matrix protein (UHS keratin).
 OS Ovis aries (sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OK NCBI_TaxID=9940;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Follicle;
 RX MEDLINE=91115951; PubMed=1703541;
 RA McKinnon P.J., Powell B.C., Rogers G.E.;
 RT "Structure and expression of genes for a class of cysteine-rich
 proteins of the cuticle layers of differentiating wool and hair
 follicles.";
 RL J. Cell Biol. 111:2587-2600(1990).

CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
 SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
 MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
 SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
 6-20 KDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
 KERATINS (40-56 kDa).
 CC -!- TISSUE SPECIFICITY: CUTICLE LAYERS OF DIFFERENTIATING
 WOOL FOLLICLES.

CC -!- DEVELOPMENTAL STAGE: AT A LATE STAGE OF FIBER DIFFERENTIATION.
 CC -!- DOMAIN: MAINLY COMPOSED OF CYS-RICH (CR), GLY-RICH (GR) AND SER-
 RICH (SR) REPEATS.
 CC -!- SIMILARITY: BELONGS TO THE UHS KERATIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; X55294; CAA39006.1; -.
 DR PIR; A36586; A36586.
 DR HSSP; P02876; 9WGA.
 KW Keratin; Repeat; Multigene family.
 SQ SEQUENCE 182 AA; 16101 MW; 9BDD4901FCB13295 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 182;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 SSCGKG 29
 |||||
 DB 48 SSCGKG 54

RESULT 35

RBPFA_MYCTU STANDARD; PRT; 183 AA.
 ID RBPFA_MYCTU
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ribosome-binding factor A.
 GN RBPFA OR RV2838C OR MT2904 OR MTCY1657.04.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1773;

OY 358 LKDGDKV 364
 |||||
 DB 120 LKDGDKV 126

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigler R., Gars S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jorgels K., Krogh A., McLean J., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; 281331; CAB03671.1; -
CC EMBL; AE007115; AAK47230.1; -
CC PIR; A70694; A70694.
CC TIGR; MT2904; -
CC TubercuList; RV2838c; -
CC HAMAP; MF_00003; -; 1.
CC InterPro; IPR000238; Rib_bind_facta.
CC Pfam; PF02033; RBFA; 1.
CC ProDom; PD007327; Rib_bind_facta; 1.
CC TIGRFAMs; TIGR00082; Ibfa; 1.
CC PROSITE; PS01319; RBFA; 1.
CC rRNA processing; Complete proteome.
CC KW RNA processing; Complete proteome; DCEPIA043CB34F92 CRC64;
CC SQ SEQUENCE 183 AA; 18998 MW; DCEPIA043CB34F92 CRC64;

```

```

Query Match 1.7%; Score 7; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 129 ADADLAR 135
DB 119 ADADLAR 125

```

```

RESULT 36
IF3_RICCN STANDARD; PRT; 185 AA.
AC 092D19;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-3.
GN INFC OR RP531.
RT

```

```

OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichelitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140 (1998).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AJ235272; CAA14980.1; -
CC PIR; B71657; B71657.
CC HSSP; P03000; ITIF.
CC HAMAP; MF_00080; -; 1.
CC InterPro; IPR001288; IF3.
CC Pfam; PF00707; IF3_C; 1.
CC Pfam; PF05198; IF3_N; 1.
CC ProDom; PD002880; IF3; 1.
CC TIGRFAMs; TIGR00168; Infc; 1.
CC PROSITE; PS00938; IF3; 1.
CC Initiation factor; Protein biosynthesis; Complete proteome.
CC KW Initiation factor; Protein biosynthesis; Complete proteome.
CC SQ SEQUENCE 185 AA; 21605 MW; EA83C963E6581BC CRC64;

```

```

Query Match 1.7%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 358 LKGDGV 364
DB 124 LKGDGV 130

```

```

RESULT 37
IF3_RICCN STANDARD; PRT; 186 AA.
AC 092HK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Translation initiation factor IF-3.
GN INFC OR RC0765.
OS Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OC NCBI_TaxId=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissendach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."

```

```

RL Science 293:2093-2098(2001).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008633; AAL0303.1; ALT_INIT.
DR HAMAP; MF_00080; -, 1.
DR InterPro; IPR001288; IF3.
DR Pfam; PF00707; IF3_C; 1.
DR Pfam; PF05198; IF3_N; 1.
DR ProDom; PD002880; IF3; 1.
DR TIGRFAMs; TIGR00168; INF3; 1.
DR PROSITE; PS00938; IF3; 1.
DR Initiation factor; Protein biosynthesis; Complete proteome.
SQ SEQUENCE 186 AA; 21665 MW; B42202AD4B293043 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 186;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LKDGDKV 364
DB 124 LKDGDKV 130

RESULT 38
Y418_VIBCH STANDARD; PRT; 187 AA.
AC 09KUT7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maf-like protein VCO418.
GN VCO418.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406633; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE MAP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004129; AAF93591.1; ALT_INIT.
DR HSSP; 002169; 1EX2.
DR TIGR; VC0418; -.
DR HAMAP; MF_00528; -, 1.
DR InterPro; IPR003697; Maf.
DR Pfam; PF02545; Maf; 1.
DR TIGRFAMs; TIGR00172; maf; 1.
KW Complete proteome.
FT ACT SITE 35
SQ SEQUENCE 187 AA; 20495 MW; DF7DA9F27667C0F7 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 187;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 LVAAEAV 145
DB 61 LVAAEAV 67

RESULT 39
PAT_ALCEFA STANDARD; PRT; 197 AA.
ID PAT_ALCEFA
AC P31668;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Phosphinothricin N-acetyltransferase (EC 2.3.1.-) (PPT N-
DE acetyltransferase) (Phosphinothricin-resistance protein).
GN PAT.
OS Alcaligenes faecalis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_TaxID=511;
RN [1]
RP SEQUENCE FROM N.A.
RA Brauer D., Bartsch K., Donn G.;
RT "Gene resistant to phosphinothricin";
RL Patent number EP0290986, 17-NOV-1988.
CC -1- FUNCTION: THIS ENZYME IS AN EFFECTOR OF PHOSPHINOTHRICIN
CC TRIPEPTIDE (PPT OR BIALAPHOS) RESISTANCE. INACTIVATES PPT BY
CC TRANSFER OF AN ACETYL GROUP.
CC -1- SIMILARITY: BELONGS TO THE ACETYLTRANSFERASE FAMILY. PAT/BAR
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A01505; -; NOT ANNOTATED - CDS.
DR EMBL; A01504; CAA00175.1; -.
DR InterPro; IPR000182; GCN5acetyltransf.
DR Pfam; PF00583; Acetyltransf; 1.
DR Transferase; Acyltransferase; Herbicide resistance;
KW Antibiotic resistance.
SQ SEQUENCE 197 AA; 21213 MW; ACA6E3B51AA55FE3 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 197;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 LNAAGDT 208
DB 188 LNAAGDT 194

RESULT 40

```

DB 180 LRRQIAE 186
 Search completed: September 8, 2003, 14:09:05
 UDB time : 32 secs

COAE BRUME
 ID COAE BRUME STANDARD; PRT; 200 AA.
 AC 08YE21;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dephospho-CoA kinase (EC 2.7.1.24) (Dephosphocoenzyme A kinase).
 OS COAE OR BMEI2057 OR BR2070.
 OS Brucella melitensis, and
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OK NCBI_TaxID=29459, 29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756588;
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujer C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyriades N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. suis; STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 CC -1- FUNCTION: Catalyzes the phosphorylation of the 3'-hydroxyl group
 CC of dephosphocoenzyme A to form coenzyme A (By similarity).
 CC -1- CATALYTIC ACTIVITY: ATP + dephospho-CoA = ADP + CoA.
 CC -1- PATHWAY: Coenzyme A (CoA) biosynthesis; fitch (last) step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the coa family.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 CC EMBL; AE009638; AAL53238.1; -;
 CC EMBL; AE014494; AAN30960.1; -;
 CC PIR; AC3509; AC3509.
 CC TIGR; BR2070; -;
 CC HAMAP; MF_00376; -; 1.
 CC InterPro; IPR001977; Depp_CoAkinase.
 CC Pfam; PF01121; Coae; 1.
 CC ProDom; PD003329; Depp_CoAkinase; 1.
 CC TIGRFAMs; TIGR00152; TIGR00152; 1.
 CC PROSITE; PS01294; COAE; FALSE NEG.
 CC Transferase; kinase; ATP-binding; Coenzyme A biosynthesis;
 CC Complete proteome.
 CC NP_BIND 8 15 ATP (POTENTIAL).
 CC SEQUENCE 200 AA; 21441 MW; B2CB67FF86C8430D CRC64;

Query Match 1.7%; Score 7; DB 1; Length 200;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:06:27 ; Search time 21 Seconds
(without alignments)
1886.738 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 412

Sequence: 1 MAFYAFKAMRAALAAVAL.....AAPQGVQTASEAKTASEAE 412

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	41.5	412	2	H81825
2	133	32.3	412	2	F81051
3	100	24.3	271	2	S42418
4	11	2.7	373	2	E85757
5	11	2.7	373	2	G90861
6	11	2.7	388	2	E87349
7	9	2.2	444	2	AH0346
8	8	1.9	107	2	C81232
9	8	1.9	111	2	E42645
10	8	1.9	111	2	D81664
11	8	1.9	122	2	D70730
12	8	1.9	129	2	C70022
13	8	1.9	171	2	D90703
14	8	1.9	171	2	G85553
15	8	1.9	171	2	B64785
16	8	1.9	181	2	A70673
17	8	1.9	183	2	A72499
18	8	1.9	183	2	G90849
19	8	1.9	183	2	F85707
20	8	1.9	290	2	T36757
21	8	1.9	296	2	C95961
22	8	1.9	306	2	H86740
23	8	1.9	348	2	C70415
24	8	1.9	408	2	AC0961
25	8	1.9	416	2	B65171
26	8	1.9	416	2	B91207
27	8	1.9	416	2	G86053
28	8	1.9	434	1	WMBE73
29	8	1.9	478	1	DPHDM2

30	8	1.9	480	2	A46702	methionyl aminope
31	8	1.9	489	2	T22088	hypothetical prote
32	8	1.9	716	2	AB1070	probable carbon st
33	8	1.9	748	2	T35818	probable integral
34	8	1.9	871	2	AH0172	membrane alanyl am
35	8	1.9	891	1	DEEC	acetaldehyde dehyd
36	8	1.9	891	2	D85704	hypothetical prote
37	8	1.9	891	2	F90846	acetaldehyde dehyd
38	8	1.9	891	2	AC0265	alcohol dehydrogen
39	8	1.9	892	2	AE0650	alcohol dehydrogen
40	8	1.9	979	2	T01566	hypothetical prote
41	8	1.9	984	2	T44496	cellulose 1,4-beta
42	8	1.9	1121	2	T02764	myosin-I binding p
43	8	1.9	1291	2	T13389	hypothetical prote
44	8	1.9	1558	2	C89114	protein C37C3.6a l
45	8	1.9	1882	2	T00069	hypothetical prote
46	8	1.9	2167	2	T34395	hypothetical prote
47	8	1.7	13	2	A38929	glutathione peroxi
48	8	1.7	79	2	S61446	cellulase (EC 3.2.
49	7	1.7	91	1	RGE8BD	mercuric resistanc
50	7	1.7	91	1	S09524	mercuric resistanc
51	7	1.7	97	2	T28947	hypothetical prote
52	7	1.7	98	2	T28629	hypothetical prote
53	7	1.7	106	2	S07520	Y4C8 protein - Rhl
54	7	1.7	106	2	A41880	host specificity p
55	7	1.7	111	2	F72054	nitrogen fixation
56	7	1.7	111	2	A86570	ribosomal protein
57	7	1.7	123	2	B81727	L24 ribosomal prot
58	7	1.7	131	2	AF3339	ribosomal protein
59	7	1.7	132	2	I39004	precocorrin-3b C17-m
60	7	1.7	134	2	A26986	cyclin-dependent k
61	7	1.7	135	2	F87515	hypothetical prote
62	7	1.7	145	2	AD2920	hypothetical prote
63	7	1.7	145	2	B37694	cofamin biosynth
64	7	1.7	147	2	H70514	hypothetical prote
65	7	1.7	151	2	F95316	probable SyxB-like
66	7	1.7	151	2	G95369	SVR82 transcrip
67	7	1.7	152	2	A87264	hypothetical prote
68	7	1.7	154	2	AE0445	biotin carboxyl ca
69	7	1.7	157	2	E97795	transacton initia
70	7	1.7	157	2	C71060	hypothetical prote
71	7	1.7	158	2	A12718	conserved hypotet
72	7	1.7	158	2	E97500	hypothetical prote
73	7	1.7	163	2	T15985	hypothetical prote
74	7	1.7	168	2	AE0138	peptidoglycan-asso
75	7	1.7	169	2	S22496	peptidylprolyl iso
76	7	1.7	169	2	T35937	hypothetical prote
77	7	1.7	172	2	S54833	peptidylprolyl iso
78	7	1.7	172	2	T10056	peptidylprolyl iso
79	7	1.7	180	2	F75607	arsenate reductase
80	7	1.7	181	2	H69157	hypothetical prote
81	7	1.7	182	2	A36686	ultra-high-sulfur
82	7	1.7	183	2	A70694	probable tRNA - My
83	7	1.7	185	2	B71657	transacton initia
84	7	1.7	187	2	B65100	hypothetical 19.4
85	7	1.7	191	2	T46412	keratin KAP5.4 - s
86	7	1.7	197	2	T46413	keratin KAP5.5 - s
87	7	1.7	200	2	AC3509	dephospho-CoA kina
88	7	1.7	202	2	B87342	hypothetical prote
89	7	1.7	203	2	T36403	probable tetr-fami
90	7	1.7	204	2	AG2383	transacton initia
91	7	1.7	205	1	OPBOE	glutathione peroxi
92	7	1.7	205	2	C82325	muf protein VC0418
93	7	1.7	206	2	T36643	probable integral
94	7	1.7	209	2	D81270	hypothetical prote
95	7	1.7	209	2	G87569	conserved hypotet
96	7	1.7	212	2	AD2356	AhpC/GSA family pr
97	7	1.7	212	2	C75049	hypothetical prote
98	7	1.7	213	2	F84532	hypothetical prote
99	7	1.7	214	2	A97624	hypothetical 23.8K
100	7	1.7	214	2	AH2846	hydrolase [importe
101	7	1.7	217	2	T06591	cellulase (EC 3.2.
102	7	1.7	223	2	T35908	probable membrane

103	7	1.7	226	2	T06378	oleosin p24, isofo
104	7	1.7	230	2	E87578	hypothetical prote
105	7	1.7	231	2	D69883	conserved hypotbet
106	7	1.7	232	2	C71908	3-oxoacid CoA tran
107	7	1.7	233	2	T43285	probable progester
108	7	1.7	240	2	AE3612	multidrug resistan
109	7	1.7	242	2	D81438	dihydroadipicollinat
110	7	1.7	242	2	B82815	conserved hypotbet
111	7	1.7	243	2	T05307	hypothetical prote
112	7	1.7	244	2	AB0195	3-oxoacyl-l-acyl-ca
113	7	1.7	249	2	E82917	methionine aminope
114	7	1.7	251	2	S39202	cellulase (EC 3.2.
115	7	1.7	260	2	C69130	shikimate 5-dehydr
116	7	1.7	267	2	E75471	transcription regu
117	7	1.7	270	2	JN0767	homobox protein H
118	7	1.7	271	2	S30230	homeotic protein H
119	7	1.7	273	2	AB2811	conserved hypotbet
120	7	1.7	273	2	E97589	hypothetical prote
121	7	1.7	274	2	S67444	probable 5'-AMP-ac
122	7	1.7	275	2	C87153	acyl-[ACPl] deatur
123	7	1.7	277	2	F84336	pyrroline-5-carbox
124	7	1.7	278	1	G69964	ribose phosphate p
125	7	1.7	281	2	AB4371	hypothetical prote
126	7	1.7	283	2	B82631	hypothetical prote
127	7	1.7	286	2	H83101	conserved hypotbet
128	7	1.7	286	2	E69516	hypothetical prote
129	7	1.7	288	2	T45715	hypothetical prote
130	7	1.7	293	2	S77419	hypothetical prote
131	7	1.7	295	2	AC1268	cell-shape determi
132	7	1.7	295	2	AE1630	cell-shape determi
133	7	1.7	302	2	A96841	hypothetical prote
134	7	1.7	308	2	F82509	conserved hypotbet
135	7	1.7	312	2	S18387	heme oxygenase - r
136	7	1.7	315	2	E81937	probable transmemb
137	7	1.7	315	2	B81168	transporter NMB070
138	7	1.7	317	2	AB0450	trehalase operon r
139	7	1.7	319	2	S61447	cellulase (EC 3.2.
140	7	1.7	324	2	S61447	cellulase (EC 3.2.
141	7	1.7	324	2	D82845	copper resistance
142	7	1.7	324	2	T23876	hypothetical prote
143	7	1.7	326	2	B40141	mitochondrial solu
144	7	1.7	330	2	F81656	conserved hypotbet
145	7	1.7	332	2	S76622	hypothetical prote
146	7	1.7	334	2	T03520	probable nicotinat
147	7	1.7	336	2	F95925	probable cell-wall
148	7	1.7	349	2	A40141	microhondrial solu
149	7	1.7	350	2	T36611	probable integral
150	7	1.7	354	2	T04779	hypothetical prote
151	7	1.7	355	2	JC1249	peroxidase (EC 1.1
152	7	1.7	355	2	F90391	conserved hypotbet
153	7	1.7	357	2	J01647	SH1 protein - hnm
154	7	1.7	358	2	C82281	ferric vibriobacti
155	7	1.7	361	2	T48297	protein phosphatas
156	7	1.7	363	2	T44150	hypothetical prote
157	7	1.7	364	2	S34355	peroxidase (EC 1.1
158	7	1.7	367	2	T34884	probable integral
159	7	1.7	367	2	E81379	probable membrane
160	7	1.7	369	2	B87645	HLV family secret
161	7	1.7	372	2	A36693	lignin peroxidase
162	7	1.7	373	2	B70673	probable ddla - My
163	7	1.7	374	2	T04778	hypothetical prote
164	7	1.7	376	2	C95878	probable dehydrog
165	7	1.7	380	2	AE3324	UDP-N-acetylglucos
166	7	1.7	380	2	D70516	hypothetical prote
167	7	1.7	380	2	T45819	villin 3 homolog F
168	7	1.7	385	2	T36688	probable serine pr
169	7	1.7	386	2	S52981	lycogene cyclase -
170	7	1.7	393	2	D81139	2-oxoglutarate deh
171	7	1.7	393	2	T35894	hypothetical prote
172	7	1.7	395	2	T35481	hypothetical prote
173	7	1.7	396	2	A70581	hypothetical prote
174	7	1.7	396	2	E83393	RND multidrug enfl
175	7	1.7	398	2	F69266	3-ketoacyl-CoA thl
176	7	1.7	402	2	T34715	probable ornithine
177	7	1.7	403	2	A81882	probable dihydroit
178	7	1.7	404	2	D70977	hypothetical prote
179	7	1.7	405	2	A36896	tetracycline resis
180	7	1.7	405	2	A32812	repa protein - Agr
181	7	1.7	406	2	G72538	hypothetical prote
182	7	1.7	409	2	T35118	probable secreted
183	7	1.7	417	2	S51961	FUN50 protein - ye
184	7	1.7	417	2	G82968	serine hydroxymeth
185	7	1.7	417	2	A75563	2-oxoglutarate deh
186	7	1.7	417	2	B83487	probable oxidoredu
187	7	1.7	418	2	C63341	serine hydroxymeth
188	7	1.7	418	2	E72776	hypothetical prote
189	7	1.7	420	2	H84360	hypothetical prote
190	7	1.7	433	2	T43961	hypothetical prote
191	7	1.7	435	2	T36780	hypothetical prote
192	7	1.7	436	2	F91127	hypothetical prote
193	7	1.7	436	2	B85972	hypothetical prote
194	7	1.7	441	2	A75619	cobyrinic acid a,c
195	7	1.7	441	2	T48960	vacuolar H(+)-ATPa
196	7	1.7	447	2	B82490	probable Na+/H+ an
197	7	1.7	447	2	G84001	glycine dehydrogen
198	7	1.7	458	2	A48392	alpha 2C4 adrenoce
199	7	1.7	458	2	A40352	alpha-2-adrenoergi
200	7	1.7	458	2	A37869	alpha-2B-adrenoergi
201	7	1.7	458	2	T49480	alpha-2 adrenoergi
202	7	1.7	458	2	T49114	hypothetical prote
203	7	1.7	470	2	S33639	finger protein esc
204	7	1.7	479	2	T05588	cellulase (EC 3.2.
205	7	1.7	479	2	T07025	cellulase (EC 3.2.
206	7	1.7	481	2	T04021	cellulase (EC 3.2.
207	7	1.7	481	2	I49072	protein kinase - m
208	7	1.7	486	2	T06770	cellulase (EC 3.2.
209	7	1.7	486	2	T07885	cellulase (EC 3.2.
210	7	1.7	489	2	E86366	protein P26F24.6 [
211	7	1.7	489	2	T06350	cellulase (EC 3.2.
212	7	1.7	492	2	E86731	endo-1,4-beta-gluc
213	7	1.7	493	2	T06060	cellulase (EC 3.2.
214	7	1.7	494	2	T06059	cellulase (EC 3.2.
215	7	1.7	494	2	S11946	cellulase (EC 3.2.
216	7	1.7	501	2	A86158	endo-1,4-beta gluc
217	7	1.7	501	2	T52135	cellulase (EC 3.2.
218	7	1.7	502	2	A10151	probable bacteriop
219	7	1.7	505	2	T07883	cellulase (EC 3.2.
220	7	1.7	507	2	JC7226	endo-1,3(4)-beta-9
221	7	1.7	507	2	S46500	cellulase (EC 3.2.
222	7	1.7	508	2	T37224	hypothetical prote
223	7	1.7	510	2	T10734	cellulase (EC 3.2.
224	7	1.7	510	2	S57808	cellulase (EC 3.2.
225	7	1.7	515	2	T46610	cellulase (EC 3.2.
226	7	1.7	516	2	T01419	cellulase (EC 3.2.
227	7	1.7	519	1	S03564	aldehyde dehydroge
228	7	1.7	520	2	D70776	probable export pr
229	7	1.7	528	2	C85056	probable DNA-bindi
230	7	1.7	530	2	A70589	probable efpa prot
231	7	1.7	533	2	S18539	actIVA-1 protein -
232	7	1.7	535	2	C87113	probable secreted
233	7	1.7	539	2	F70737	chaperonin groEL1
234	7	1.7	539	2	T50579	probable membrane
235	7	1.7	541	2	E86789	protein T23E18.10
236	7	1.7	542	2	S73749	transport system p
237	7	1.7	543	2	T34681	probable secreted
238	7	1.7	544	2	S58532	mask protein (trnx
239	7	1.7	551	2	R83015	hypothetical prote
240	7	1.7	559	2	B84213	hypothetical prote
241	7	1.7	568	1	S05532	gamma-glutamyltran
242	7	1.7	571	2	T01511	hypothetical prote
243	7	1.7	574	2	B83388	probable carbamoyl
244	7	1.7	582	2	B70389	translation elonga
245	7	1.7	584	2	D84264	hypothetical prote
246	7	1.7	593	2	T20650	hypothetical prote
247	7	1.7	597	1	S32039	chitinase (EC 3.2.
248	7	1.7	597	2	T25547	hypothetical prote

249	7	1.7	600	2	D83286	hypothetical prote
250	7	1.7	608	2	A64992	sensor protein Ato
251	7	1.7	610	2	T35797	secreted chitinase
252	7	1.7	616	2	B3586	C4-dicarboxylate t
253	7	1.7	621	2	D96032	C4-dicarboxylate t
254	7	1.7	622	2	T49426	Type 2C protein Ph
255	7	1.7	622	2	B83386	hypothetical prote
256	7	1.7	627	2	B96527	protein F27015.28
257	7	1.7	631	2	T35234	probable secreted
258	7	1.7	631	2	T38167	electron transfer
259	7	1.7	640	2	A41726	homeotic protein B
260	7	1.7	642	2	S27806	homeotic protein B
261	7	1.7	645	2	H70783	hypothetical prote
262	7	1.7	649	2	T40646	probable protein i
263	7	1.7	659	2	E84176	DNA mismatch repair
264	7	1.7	677	2	T45682	hypothetical prote
265	7	1.7	681	2	S73550	DNA polymerase III
266	7	1.7	687	2	B81027	glycyl-tRNA synth
267	7	1.7	687	2	G81970	probable glycine-t
268	7	1.7	689	1	SYECGB	glycine-tRNA ligas
269	7	1.7	689	2	B91184	glycine tRNA synth
270	7	1.7	689	2	G86030	glycine tRNA synth
271	7	1.7	689	2	A80494	glycine-tRNA ligas
272	7	1.7	712	2	T27165	hypothetical prote
273	7	1.7	721	2	S56580	carbon starvation
274	7	1.7	721	2	A98293	probable carbon st
275	7	1.7	721	2	D86134	1-deoxy-D-xyulose
276	7	1.7	735	2	T08140	hypothetical prote
277	7	1.7	762	2	T14815	hypothetical prote
278	7	1.7	774	2	D83208	probable ferredoxi
279	7	1.7	789	2	A82688	Na+/H+ antiporter
280	7	1.7	789	2	H97469	probable NADH dehy
281	7	1.7	878	2	H83184	phosphoenolpyruvat
282	7	1.7	887	2	S57430	lipo protein - huma
283	7	1.7	896	2	S57723	DNA polymerase B1
284	7	1.7	901	2	B84210	probable helicase
285	7	1.7	906	2	G70767	glycoprotein 130 -
286	7	1.7	917	2	I49699	interleukin-6 sign
287	7	1.7	918	2	A44257	hypothetical prote
288	7	1.7	962	2	T22459	villin 3 [imported
289	7	1.7	966	2	T50668	sarcosine oxidase
290	7	1.7	1005	2	B82969	cellulose
291	7	1.7	1062	2	H83966	cellulase (EC 3.2.
292	7	1.7	1070	2	S75712	middle wall protei
293	7	1.7	1084	2	A28555	S-layer protein pr
294	7	1.7	1099	2	T14850	adenylate cyclase
295	7	1.7	1144	1	A39833	delta endotoxin -
296	7	1.7	1186	2	T18210	urea amidolyase-re
297	7	1.7	1207	2	H87475	hemeamidase-sensi
298	7	1.7	1212	2	A57187	hypothetical prote
299	7	1.7	1230	2	T22458	probable SpoIIIF-f
300	7	1.7	1345	2	A87102	collagen alpha 1(I
301	7	1.7	1466	1	C8H07L	glutamate synthase
302	7	1.7	1530	2	B82085	microtubule-associ
303	7	1.7	1734	2	A54602	probable tape-meas
304	7	1.7	1787	2	AG1360	probable tape-meas
305	7	1.7	1788	2	AH1447	extracellular matr
306	7	1.7	2055	2	T31110	hypothetical prote
307	7	1.7	2946	2	T00867	hypothetical prote
308	7	1.7	2957	2	T33152	hypothetical prote
309	7	1.7	3079	1	RGBY12	legumen GTPase-ac
310	7	1.7	3436	2	S55659	legumen protein 6
311	7	1.7	3535	2	B83641	probable hemagglu
312	7	1.7	3624	2	AD0835	large repetitive p
313	7	1.7	5627	2	C83339	hypothetical prote
314	7	1.7	5627	2	A60722	cryptic fibrillar p
315	6	1.5	40	1	PDF18G	antifreeze protein
316	6	1.5	45	2	A05163	antifreeze protein
317	6	1.5	53	2	PC4101	iron-sulfur protei
318	6	1.5	55	2	C82796	hypothetical prote
319	6	1.5	57	2	A61381	prothymosin alpha
320	6	1.5	59	2	A71845	hypothetical prote
321	6	1.5	60	2	S13476	ferredoxin 2[4Fe-4
322	6	1.5	60	2	A29667	pulmonary surfacta
323	6	1.5	61	2	A13331	cbim protein [impo
324	6	1.5	62	2	AD2476	hypothetical prote
325	6	1.5	63	2	S03864	hypothetical prote
326	6	1.5	65	2	H64324	ferredoxin 2[4Fe-4
327	6	1.5	65	2	A95330	ferredoxin 2[4Fe-4
328	6	1.5	65	2	C72153	hypothetical prote
329	6	1.5	66	2	C72153	hypothetical prote
330	6	1.5	66	2	A36839	PIL protein - vari
331	6	1.5	66	2	T28459	CIL protein - vari
332	6	1.5	68	2	UC7133	hypothetical prote
333	6	1.5	69	2	S60826	antioxidant protei
334	6	1.5	71	2	H97137	M protein precurs
335	6	1.5	72	2	A35731	hypothetical prote
336	6	1.5	72	2	S68883	Ca2+-transporting
337	6	1.5	72	2	S74974	light-harvesting p
338	6	1.5	72	2	C71355	hypothetical prote
339	6	1.5	73	2	C71355	probable ribosomal
340	6	1.5	73	2	H81298	probable molybdopt
341	6	1.5	73	2	C82586	hypothetical prote
342	6	1.5	77	2	C95909	conserved hypotnet
343	6	1.5	78	2	D81267	50S ribosomal prot
344	6	1.5	78	2	S74715	ribosomal protein
345	6	1.5	79	2	AG2134	50S ribosomal prot
346	6	1.5	80	2	T36418	hypothetical prote
347	6	1.5	80	2	OSH07B	cytochrome-c oxida
348	6	1.5	80	2	E90686	hypothetical prote
349	6	1.5	80	2	A70688	hypothetical prote
350	6	1.5	81	2	A85537	unknown [imported]
351	6	1.5	81	2	H64814	molybdopterin bios
352	6	1.5	81	2	A82251	molybdopterin bios
353	6	1.5	81	2	F90736	molybdopterin bios
354	6	1.5	81	2	C80558	molybdopterin bios
355	6	1.5	83	2	AC0598	molybdopterin conv
356	6	1.5	83	2	D87676	hypothetical prote
357	6	1.5	83	2	T34904	hypothetical prote
358	6	1.5	83	2	T02136	hypothetical prote
359	6	1.5	84	2	T07115	hypothetical prote
360	6	1.5	84	2	D70672	metallothionein-11
361	6	1.5	85	2	S62832	hypothetical prote
362	6	1.5	85	2	OSB07B	cytochrome-c oxida
363	6	1.5	88	1	C97819	hypothetical prote
364	6	1.5	92	2	D85805	hypothetical prote
365	6	1.5	93	2	C75478	unknown protein en
366	6	1.5	95	2	C97819	hypothetical prote
367	6	1.5	96	2	A86912	conserved hypotnet
368	6	1.5	96	2	AG3001	hypothetical prote
369	6	1.5	98	1	AF3827	50S ribosomal prot
370	6	1.5	98	1	AF3827	chaperonin groES -
371	6	1.5	98	1	AF3827	10K chaperonin gro
372	6	1.5	100	2	E86121	cytochrome b (562)
373	6	1.5	100	2	E91280	hypothetical prote
374	6	1.5	100	2	I40669	hypothetical prote
375	6	1.5	100	2	UQ0205	hypothetical 12.5K
376	6	1.5	101	2	H86885	50S ribosomal prot
377	6	1.5	102	2	C87282	conserved hypotnet
378	6	1.5	102	2	T35134	hypothetical prote
379	6	1.5	104	2	S75246	hypothetical prote
380	6	1.5	104	2	C82935	hypothetical prote
381	6	1.5	105	2	F72662	conserved hypotnet
382	6	1.5	105	2	G87591	hypothetical prote
383	6	1.5	106	2	S13027	hypothetical prote
384	6	1.5	106	2	S37865	cysteine proteinas
385	6	1.5	106	2	C83584	hypothetical prote
386	6	1.5	107	1	PVPK2	conserved hypotnet
387	6	1.5	107	2	A38375	pavabunin beta (
388	6	1.5	107	2	B70650	oryzaacystatin II -
389	6	1.5	107	2	AG3321	probable emtr prot
390	6	1.5	107	2	C81080	hypothetical prote
391	6	1.5	107	2	AF1965	hypothetical prote
392	6	1.5	108	2	I64217	hypothetical prote
393	6	1.5	108	2	S34941	ribosomal protein
394	6	1.5	108	2	T40960	ribosomal protein S
395	6	1.5	108	2	A48831	hypothetical prote
396	6	1.5	108	2	A48831	vitelline membrane

395	6	1.5	108	2	F72477	468	6	1.5	135	2	B49218	hemagglutinin homo
396	6	1.5	108	2	A95904	469	6	1.5	135	2	H72499	hypothetical prote
397	6	1.5	109	1	TNBOA1	470	6	1.5	135	2	G87686	hypothetical prote
398	6	1.5	109	2	T43705	471	6	1.5	135	2	JC4734	flagellar protein
399	6	1.5	109	2	G64609	472	6	1.5	135	2	T48927	hypothetical prote
400	6	1.5	110	1	PEBO	473	6	1.5	135	2	S67924	spore-wall fungal
401	6	1.5	110	2	B55863	474	6	1.5	136	2	C81226	hypothetical prote
402	6	1.5	110	2	T45387	475	6	1.5	136	2	B82787	hypothetical prote
403	6	1.5	111	2	S62830	476	6	1.5	137	2	S76236	hypothetical prote
404	6	1.5	112	2	D70535	477	6	1.5	137	2	E72614	hypothetical prote
405	6	1.5	113	2	S62528	478	6	1.5	138	1	BMD08	development-specif
406	6	1.5	113	2	S23439	479	6	1.5	138	2	T07203	hypothetical prote
407	6	1.5	113	2	JN0646	480	6	1.5	139	2	D70028	conserved hypotnet
408	6	1.5	113	2	G64945	481	6	1.5	139	2	C70718	hypothetical prote
409	6	1.5	113	2	E90947	482	6	1.5	139	2	C87544	hypothetical prote
410	6	1.5	113	2	A85796	483	6	1.5	140	2	T49481	hypothetical prote
411	6	1.5	113	2	AH0397	484	6	1.5	140	2	AD2423	hypothetical prote
412	6	1.5	113	2	T45276	485	6	1.5	141	2	C69007	ribosomal protein
413	6	1.5	114	2	G81353	486	6	1.5	141	2	T34890	probable IS1648 tr
414	6	1.5	115	2	B75535	487	6	1.5	142	2	C32223	conserved hypotnet
415	6	1.5	116	2	T37186	488	6	1.5	142	2	G72050	transcription regu
416	6	1.5	116	2	B75456	489	6	1.5	142	2	D86574	conserved hypotnet
417	6	1.5	116	2	S52219	490	6	1.5	143	2	AC2693	hypothetical prote
418	6	1.5	116	2	D84285	491	6	1.5	144	2	DB3152	hypothetical prote
419	6	1.5	117	2	F95862	492	6	1.5	144	2	C96004	conserved hypotnet
420	6	1.5	118	2	F87447	493	6	1.5	144	2	B70949	hypothetical prote
421	6	1.5	119	1	CUPSAM	494	6	1.5	144	2	G64423	hypothetical prote
422	6	1.5	119	2	S17123	495	6	1.5	145	2	F69214	heat shock protein
423	6	1.5	119	2	C90864	496	6	1.5	145	2	C84610	probable beta-hydr
424	6	1.5	119	2	F85754	497	6	1.5	145	2	F83267	conserved hypotnet
425	6	1.5	119	2	C82795	498	6	1.5	146	1	HBAK	hemoglobin beta ch
426	6	1.5	119	2	H97183	499	6	1.5	146	1	HBAQ	hemoglobin beta ch
427	6	1.5	120	1	H64900	500	6	1.5	146	2	G7474	probable asnc-fami
428	6	1.5	120	2	T35920	501	6	1.5	146	2	T10511	hypothetical prote
429	6	1.5	121	2	T03318	502	6	1.5	146	2	AE3630	transcription regu
430	6	1.5	121	2	S34233	503	6	1.5	147	2	C31844	K118 protein - Str
431	6	1.5	121	2	F83940	504	6	1.5	148	2	D75276	response regulator
432	6	1.5	122	2	D42645	505	6	1.5	148	2	AC1122	hypothetical prote
433	6	1.5	122	2	B86570	506	6	1.5	149	2	D53399	protein (imported
434	6	1.5	122	2	E81664	507	6	1.5	150	2	G87590	cytochrome c-type
435	6	1.5	122	2	G72054	508	6	1.5	150	2	A70305	(3R)-hydroxymyrist
436	6	1.5	123	2	G97885	509	6	1.5	150	2	S58172	mlthramycin polyke
437	6	1.5	124	2	B70148	510	6	1.5	150	2	D87652	hypothetical prote
438	6	1.5	124	2	F71347	511	6	1.5	151	2	A24475	superoxide dismuta
439	6	1.5	124	2	T05467	512	6	1.5	151	2	AC2084	phosphonate metabo
440	6	1.5	125	2	T50866	513	6	1.5	151	2	T15428	hypothetical prote
441	6	1.5	125	2	T34695	514	6	1.5	152	2	A83053	conserved hypotnet
442	6	1.5	125	2	AF1932	515	6	1.5	152	2	A87679	conserved hypotnet
443	6	1.5	126	2	T21448	516	6	1.5	153	2	S21767	hypothetical prote
444	6	1.5	126	2	H70634	517	6	1.5	153	2	F69267	cytochrome-c oxida
445	6	1.5	126	2	PN0542	518	6	1.5	154	2	B83304	hypothetical prote
446	6	1.5	127	2	B64034	519	6	1.5	154	2	AC3461	hypothetical prote
447	6	1.5	127	2	B75353	520	6	1.5	154	2	D97415	hypothetical prote
448	6	1.5	127	2	A72712	521	6	1.5	155	2	C64921	outer membrane lip
449	6	1.5	128	1	CBE662	522	6	1.5	155	2	P90922	outer membrane lip
450	6	1.5	128	2	AF1056	523	6	1.5	155	2	B85771	probable outer mem
451	6	1.5	128	2	B70647	524	6	1.5	155	2	A10693	outer membrane lip
452	6	1.5	128	2	T50723	525	6	1.5	155	2	H86972	riboflavin synthas
453	6	1.5	128	2	S72600	526	6	1.5	155	2	D87399	cytochrome c fam1
454	6	1.5	129	2	T16369	527	6	1.5	156	2	A97397	deoxyuridine 5'-tr
455	6	1.5	130	2	T34729	528	6	1.5	156	2	AB2615	deoxyuridine 5'-tr
456	6	1.5	131	2	B72523	529	6	1.5	156	2	AH1606	acetyl-CoA carboxy
457	6	1.5	131	2	AB1525	530	6	1.5	156	2	C97503	hypothetical prote
458	6	1.5	131	2	A87355	531	6	1.5	156	2	S08427	12k protein ma12A
459	6	1.5	131	2	S18540	532	6	1.5	156	2	AB0079	hypothetical prote
460	6	1.5	132	2	G82423	533	6	1.5	157	2	F91172	probable phosphor
461	6	1.5	133	2	H87681	534	6	1.5	157	2	F66018	hypothetical prote
462	6	1.5	134	1	F71142	535	6	1.5	157	2	AB2539	hypothetical prote
463	6	1.5	134	2	AC1166	536	6	1.5	158	2	T06078	hypothetical prote
464	6	1.5	134	2	B83943	537	6	1.5	158	2	B91251	hypothetical prote
465	6	1.5	134	2	D97582	538	6	1.5	158	2	H85620	hypothetical prote
466	6	1.5	134	2	AD2803	539	6	1.5	159	2	H75391	hypothetical prote
467	6	1.5	135	2	D84730	540	6	1.5	160	2	F75450	hypothetical prote

541	6	1.5	160	2	C83894	hypothetical prote
542	6	1.5	160	2	P95334	Nex18 Symbiocell
543	6	1.5	160	2	AE0577	conserved hypothet
544	6	1.5	160	2	AB1948	hypothetical prote
545	6	1.5	160	2	AC2459	transposase alr522
546	6	1.5	161	1	C69405	hypothetical prote
547	6	1.5	161	1	D97145	molYdenum cofacto
548	6	1.5	161	2	F87277	hypothetical prote
549	6	1.5	161	2	F64038	hypothetical prote
550	6	1.5	162	2	C40655	probable replicat
551	6	1.5	162	2	F64553	hypothetical prote
552	6	1.5	162	2	A75359	hypothetical prote
553	6	1.5	162	2	H70927	probable rpsp prot
554	6	1.5	162	2	AB1950	hypothetical prote
555	6	1.5	163	2	T47812	hypothetical prote
556	6	1.5	163	2	G71180	probable 3-isoprop
557	6	1.5	164	1	TVBE11	transforming prote
558	6	1.5	165	2	A45192	ribosomal protein
559	6	1.5	165	2	G81663	thiol peroxidase P
560	6	1.5	165	2	B83328	hypothetical prote
561	6	1.5	165	2	B70688	hypothetical prote
562	6	1.5	165	2	S15800	hypothetical prote
563	6	1.5	166	2	T07949	calcium binding pr
564	6	1.5	166	2	D24886	P(-)rp(+) fibrinai
565	6	1.5	167	2	E70396	hypothetical prote
566	6	1.5	169	2	T35087	probable transcript
567	6	1.5	169	2	H81980	probable periplasm
568	6	1.5	169	2	B81036	thiorodoxin NMB184
569	6	1.5	169	2	AG3197	ECF family sigma f
570	6	1.5	170	2	I40612	3'-aminoglycoside
571	6	1.5	170	2	S57935	CocA protein precu
572	6	1.5	171	2	G87218	probable cytidine/
573	6	1.5	171	2	AH2089	(3r)-hydroxymyrist
574	6	1.5	171	2	S15911	hypothetical prote
575	6	1.5	172	2	T06073	peptidylprolyl iso
576	6	1.5	172	2	UN0732	hypothetical 19.4k
577	6	1.5	172	2	C47119	spore coat protein
578	6	1.5	172	2	T14878	hypothetical prote
579	6	1.5	173	2	S10199	NADH2 dehydrogenas
580	6	1.5	173	2	S44414	NADH2 dehydrogenas
581	6	1.5	173	2	C97287	transcription anti
582	6	1.5	173	2	A83965	transposase (21) B
583	6	1.5	173	2	AB3648	flagellar basal-bo
584	6	1.5	174	2	A35383	superoxide dismuta
585	6	1.5	174	2	AT0651	hypothetical prote
586	6	1.5	175	1	VGMRTS	glycoprotein NCVP5
587	6	1.5	175	1	AB3051	conserved hypothet
588	6	1.5	175	2	G98234	hypothetical prote
589	6	1.5	176	2	B70445	heat shock protein
590	6	1.5	177	2	A32223	interleukin-7 prec
591	6	1.5	177	2	A69938	conserved hypothet
592	6	1.5	177	2	C95423	probable stress-in
593	6	1.5	178	2	C86885	30S ribosomal prot
594	6	1.5	178	2	C81785	conserved hypothet
595	6	1.5	178	2	H81207	conserved hypothet
596	6	1.5	178	2	S73092	hypothetical prote
597	6	1.5	178	2	T28045	hypothetical prote
598	6	1.5	178	2	AF2721	conserved hypothet
599	6	1.5	178	2	A84065	NADH-dependent FMN
600	6	1.5	179	2	H72697	hypothetical prote
601	6	1.5	179	2	T22734	hypothetical prote
602	6	1.5	180	2	E64834	probable type 1 fi
603	6	1.5	180	2	S15327	fibrinai protein M
604	6	1.5	181	2	AC0505	fibrinai chain (im
605	6	1.5	181	2	A83291	5-formyltetrahydro
606	6	1.5	181	2	C84592	hypothetical prote
607	6	1.5	181	2	T45990	hypothetical prote
608	6	1.5	181	2	B70170	hypothetical prote
609	6	1.5	181	2	AE0437	probable lipoprote
610	6	1.5	182	1	YOE0CT1	type 1 fibrinai pr
611	6	1.5	182	2	AH0227	Cbdiaacylglycerol-
612	6	1.5	182	2	A91288	major type 1 subun
613	6	1.5	182	2	D86129	major type 1 subun

614	6	1.5	182	2	T04682	senescence-associ
615	6	1.5	182	2	T17505	hypothetical prote
616	6	1.5	183	2	T07241	hypothetical prote
617	6	1.5	183	2	E97413	hypothetical prote
618	6	1.5	183	2	C90212	conserved hypothet
619	6	1.5	184	2	T35836	probable pyrroline
620	6	1.5	184	2	S65821	RNA-directed DNA p
621	6	1.5	185	2	D90666	probable CI repres
622	6	1.5	185	2	H85516	conserved regulator
623	6	1.5	185	2	AD2661	conserved hypothet
624	6	1.5	185	2	B97443	hypothetical prote
625	6	1.5	185	2	S26602	hypothetical prote
626	6	1.5	185	2	A82636	conserved hypothet
627	6	1.5	185	2	AB2241	hypothetical prote
628	6	1.5	185	2	D85040	hypothetical prote
629	6	1.5	186	1	H69309	L-fucose-phospha
630	6	1.5	186	1	IVB0R2	interferon beta-2
631	6	1.5	186	1	IVB0R3	interferon beta-3
632	6	1.5	186	2	B83322	CDE-diacylglycerol
633	6	1.5	186	2	D81076	conserved hypothet
634	6	1.5	186	2	T50403	probable succinate
635	6	1.5	186	2	AC2743	conserved hypothet
636	6	1.5	186	2	B97524	hypothetical prote
637	6	1.5	187	2	H81347	hypothetical prote
638	6	1.5	187	2	F81836	probable decarboxy
639	6	1.5	187	2	D81130	conserved hypothet
640	6	1.5	187	2	B87303	chemotaxis protein
641	6	1.5	187	2	E72390	orotate phosphorib
642	6	1.5	187	2	T35619	hypothetical prote
643	6	1.5	188	2	S49192	GCR 1 protein - fr
644	6	1.5	188	2	S11490	coat protein - cac
645	6	1.5	188	2	T02822	probable membrane
646	6	1.5	188	2	A75382	hypothetical prote
647	6	1.5	189	2	B42514	H2R protein - vacc
648	6	1.5	189	2	S70269	outer surface prot
649	6	1.5	189	2	S70263	outer surface prot
650	6	1.5	189	2	S27171	NADH2 dehydrogenas
651	6	1.5	189	2	C89992	hypothetical prote
652	6	1.5	190	2	A47617	site-specific reco
653	6	1.5	190	2	F75423	proteinnase I - Dei
654	6	1.5	190	2	S70261	outer surface prot
655	6	1.5	190	2	S39485	leaf-specific prot
656	6	1.5	191	2	AC0541	probable fibrinai
657	6	1.5	191	2	AC0955	probable fibrinai
658	6	1.5	191	2	G81808	hypothetical prote
659	6	1.5	191	2	T18822	hypothetical prote
660	6	1.5	192	2	D70303	imdaazoleglycerolp
661	6	1.5	192	2	S70272	outer surface prot
662	6	1.5	193	1	H1BP44	outer surface prot
663	6	1.5	193	2	S70260	outer surface prot
664	6	1.5	193	2	S70264	outer surface prot
665	6	1.5	193	2	S70265	outer surface prot
666	6	1.5	193	2	T49977	hypothetical prote
667	6	1.5	193	2	AH0440	FMN reductase (EC
668	6	1.5	194	2	AF3408	CBDiaacylglycerol-
669	6	1.5	194	2	E97141	uncharacterized pr
670	6	1.5	194	2	G64033	hypothetical prote
671	6	1.5	194	2	T29438	hypothetical prote
672	6	1.5	195	2	C75254	3-octadrenyl-4-hyd
673	6	1.5	195	2	E82062	hypothetical prote
674	6	1.5	195	2	A95350	Protein (imported
675	6	1.5	195	2	T05031	cyclic phosphodies
676	6	1.5	196	2	G72718	probable transcript
677	6	1.5	197	2	S55621	hypothetical prote
678	6	1.5	197	2	T16478	hypothetical prote
679	6	1.5	198	2	AH0778	probable membrane
680	6	1.5	199	2	S54197	outer surface prot
681	6	1.5	199	2	T36682	outer surface prot
682	6	1.5	200	2	S54198	outer surface prot
683	6	1.5	200	2	T40122	outer surface prot
684	6	1.5	200	2	G84033	glutamate racemase
685	6	1.5	201	1	T29447	probable bactiraci
686	6	1.5	201	2	T07267	endopeptidase Ctp

687	6	1.5	201	2	T06679	6	760	1.5	213	2	T17334	6	hypothetical prote
688	6	1.5	201	2	S55000	6	761	1.5	213	2	AF2664	6	ATP synthase B' ch
689	6	1.5	201	2	G02310	6	762	1.5	214	2	D4681	6	glutathione transf
690	6	1.5	201	2	G70538	6	763	1.5	214	2	F70784	6	hypothetical prote
691	6	1.5	201	2	G71428	6	764	1.5	214	2	T20056	6	hypothetical prote
692	6	1.5	201	2	B72739	6	765	1.5	215	2	C46681	6	glutathione transf
693	6	1.5	202	2	H82191	6	766	1.5	215	2	T36281	6	probable hydrolase
694	6	1.5	202	2	T23654	6	767	1.5	215	2	T44346	6	hypothetical prote
695	6	1.5	202	2	AC2714	6	768	1.5	215	2	F71368	6	hypothetical prote
696	6	1.5	202	2	H97495	6	769	1.5	215	2	D84382	6	hypothetical prote
697	6	1.5	202	2	A87636	6	770	1.5	215	2	E97446	6	ATP synthase chain
698	6	1.5	202	2	U00138	6	771	1.5	216	2	G90164	6	uracil phosphorib
699	6	1.5	202	2	G70752	6	772	1.5	217	2	F72293	6	5-methylthiodenos
700	6	1.5	203	2	I40111	6	773	1.5	217	2	F71953	6	hypothetical prote
701	6	1.5	203	2	H70522	6	774	1.5	217	2	C96748	6	unknown protein T1
702	6	1.5	204	2	B31227	6	775	1.5	217	2	T34650	6	probable transfera
703	6	1.5	204	2	E95345	6	776	1.5	218	2	AB0254	6	probable fumarylac
704	6	1.5	204	2	I40112	6	777	1.5	218	2	FE1111	6	nickel-dependent h
705	6	1.5	204	2	I50480	6	778	1.5	218	2	B81906	6	probable membrane
706	6	1.5	205	2	B87296	6	779	1.5	219	2	B29224	6	GTP-binding protei
707	6	1.5	205	2	I40105	6	780	1.5	219	2	T50020	6	(3R)-hydroxymyrist
708	6	1.5	205	2	E84334	6	781	1.5	219	2	AE0724	6	probable hydrolase
709	6	1.5	205	2	C83521	6	782	1.5	220	2	AE2535	6	transcription regu
710	6	1.5	205	2	AB3600	6	783	1.5	220	2	F70906	6	probable O-methyl
711	6	1.5	206	2	AG2815	6	784	1.5	220	2	FE4327	6	H+-transporting tw
712	6	1.5	206	2	B97594	6	785	1.5	221	2	AG1595	6	hypothetical prote
713	6	1.5	206	2	AI2123	6	786	1.5	221	2	AG1961	6	weakly oligopeptid
714	6	1.5	206	2	T27766	6	787	1.5	222	2	AG0164	6	arginine transport
715	6	1.5	206	2	AE2222	6	788	1.5	222	2	AE0607	6	arginine transpor
716	6	1.5	207	2	S31011	6	789	1.5	223	2	G64158	6	butyrate-acetoacet
717	6	1.5	207	2	E97825	6	790	1.5	223	2	D70526	6	hypothetical prote
718	6	1.5	207	2	I40271	6	791	1.5	224	1	B40850	6	secretory protein
719	6	1.5	207	2	I40276	6	792	1.5	224	2	D84345	6	phosphoribosylform
720	6	1.5	207	2	S68919	6	793	1.5	224	2	A30528	6	serum amyloid P-co
721	6	1.5	207	2	S68924	6	794	1.5	224	2	D70665	6	probable urease ac
722	6	1.5	207	2	S37727	6	795	1.5	224	2	AB1377	6	ABC transporter, p
723	6	1.5	207	2	B75294	6	796	1.5	224	2	AD1746	6	ABC transporter, p
724	6	1.5	207	2	B70708	6	797	1.5	224	2	A95394	6	hypothetical prote
725	6	1.5	208	2	T48860	6	798	1.5	224	2	G86148	6	TIN6.14 protein -
726	6	1.5	208	2	F75427	6	799	1.5	225	2	AE0610	6	conserved hypothet
727	6	1.5	208	2	T36390	6	800	1.5	225	2	T50707	6	urase accessory p
728	6	1.5	208	2	AB3406	6	801	1.5	225	2	A75257	6	probable DNA-3-met
729	6	1.5	209	1	XUFP11	6	802	1.5	226	1	E70034	6	beta-phosphoglucom
730	6	1.5	209	2	I40281	6	803	1.5	226	2	AE0967	6	conserved hypothet
731	6	1.5	209	2	S72929	6	804	1.5	226	2	T14675	6	hypothetical prote
732	6	1.5	209	2	B95116	6	805	1.5	226	2	E70501	6	probable transmemb
733	6	1.5	209	2	H97985	6	806	1.5	227	2	S35735	6	DNA gyrase chain A
734	6	1.5	210	2	E69027	6	807	1.5	227	2	S94133	6	photosystem I chai
735	6	1.5	210	2	F71308	6	808	1.5	227	2	H86940	6	probable membrane
736	6	1.5	210	2	AC0667	6	809	1.5	227	2	G75269	6	conserved hypothet
737	6	1.5	210	2	G95858	6	810	1.5	227	2	S56941	6	heat shock protein
738	6	1.5	210	2	D89825	6	811	1.5	228	2	T51147	6	hypothetical prote
739	6	1.5	210	2	D84238	6	812	1.5	228	2	C70623	6	hypothetical prote
740	6	1.5	210	2	S69920	6	813	1.5	228	2	F70624	6	probable ISI560 tr
741	6	1.5	210	2	S69925	6	814	1.5	228	2	T24592	6	hypothetical prote
742	6	1.5	210	2	FA8211	6	815	1.5	229	2	S72165	6	hypothetical prote
743	6	1.5	210	2	C69652	6	816	1.5	229	2	C43330	6	gene 7 protein - p
744	6	1.5	210	2	I58391	6	817	1.5	229	2	C68809	6	hypothetical prote
745	6	1.5	211	2	S69929	6	818	1.5	229	2	F75482	6	transcription regu
746	6	1.5	211	2	D69413	6	819	1.5	230	1	UVFUS	6	cutinase (EC 3.1.-
747	6	1.5	212	2	E82061	6	820	1.5	230	1	T36672	6	membrane-spanning
748	6	1.5	212	2	T05721	6	821	1.5	231	1	RDNCFP	6	ubiquinol-cytochro
749	6	1.5	212	2	T10230	6	822	1.5	231	2	PC1108	6	recombination prot
750	6	1.5	212	2	S44805	6	823	1.5	231	2	AD1785	6	two components res
751	6	1.5	212	2	S42396	6	824	1.5	231	2	AF3037	6	transcription regu
752	6	1.5	212	2	H87018	6	825	1.5	232	2	T36726	6	probable regulator
753	6	1.5	212	2	S23934	6	826	1.5	232	2	T14939	6	hypothetical prote
754	6	1.5	212	2	AC3048	6	827	1.5	232	2	D70537	6	hypothetical prote
755	6	1.5	212	2	H98237	6	828	1.5	233	1	B69513	6	conserved hypothet
756	6	1.5	213	2	S26035	6	829	1.5	233	2	B24796	6	glyceraldehyde-3-p
757	6	1.5	213	2	T02871	6	830	1.5	233	2	C75290	6	hypothetical prote
758	6	1.5	213	2	E75855	6	831	1.5	233	2	S57924	6	macrophage infecti
759	6	1.5	213	2	B70707	6	832	1.5	234	1	FE5080	6	hypothetical prote

833	6	1.5	234	2	C91107	906	6	1.5	249	2	C84185	hypothetical prote
834	6	1.5	234	2	F85952	907	6	1.5	249	2	B87309	hypothetical prote
835	6	1.5	234	2	AD0878	908	6	1.5	249	2	B87575	ABC transporter. A
836	6	1.5	234	2	B83065	909	6	1.5	249	2	B87597	transcription regu
837	6	1.5	234	2	154074	910	6	1.5	249	2	B87375	hypothetical prote
838	6	1.5	234	2	C87713	911	6	1.5	249	2	B98248	hypothetical prote
839	6	1.5	234	2	155448	912	6	1.5	250	2	AF1636	N-acetylglutamate
840	6	1.5	234	2	B69037	913	6	1.5	250	2	H90400	hypothetical prote
841	6	1.5	235	2	H70556	914	6	1.5	250	2	H86637	conserved hypotnet
842	6	1.5	235	2	S39652	915	6	1.5	250	2	JC7339	maleate isomerase
843	6	1.5	235	2	A99186	916	6	1.5	251	2	S28171	H-transporing tw
844	6	1.5	235	2	A13100	917	6	1.5	251	2	E90206	hypothetical prote
845	6	1.5	235	2	A12613	918	6	1.5	251	2	D90243	conserved hypotnet
846	6	1.5	236	1	T46967	919	6	1.5	251	2	T03535	precocorrin-6x redic
847	6	1.5	236	1	D71376	920	6	1.5	251	2	D82667	hypothetical prote
848	6	1.5	236	2	H95939	921	6	1.5	252	2	C83877	transcription regu
849	6	1.5	236	2	AH3247	922	6	1.5	252	2	S50806	transcription regu
850	6	1.5	237	2	AE2878	923	6	1.5	253	2	F83134	hypothetical prote
851	6	1.5	237	2	UC4914	924	6	1.5	253	2	C72758	probable short-cha
852	6	1.5	237	2	F95271	925	6	1.5	253	2	F87371	transcription regu
853	6	1.5	237	2	AG3573	926	6	1.5	253	2	A38201	transcription fact
854	6	1.5	237	2	AE2273	927	6	1.5	253	2	AG2627	hypothetical prote
855	6	1.5	237	2	A72677	928	6	1.5	253	2	A82639	hypothetical prote
856	6	1.5	238	2	T49297	929	6	1.5	253	2	H81286	probable sugar nuc
857	6	1.5	238	2	D82619	930	6	1.5	254	1	BVBV22	PT122 protein pre
858	6	1.5	238	2	A44343	931	6	1.5	254	2	D90929	probable transport
859	6	1.5	239	2	T51534	932	6	1.5	254	2	S34724	probable oxidoredu
860	6	1.5	239	2	G97654	933	6	1.5	254	2	A83345	probable thioester
861	6	1.5	239	2	D87494	934	6	1.5	254	2	A97263	lactose phosphotra
862	6	1.5	239	2	T36572	935	6	1.5	254	2	T36161	hypothetical prote
863	6	1.5	240	2	E72170	936	6	1.5	254	2	H85777	probable transport
864	6	1.5	240	2	SE1872	937	6	1.5	255	2	T23488	hypothetical prote
865	6	1.5	241	1	QOCWV2	938	6	1.5	255	2	T03519	cobalamin (5'-phos
866	6	1.5	241	2	G97654	939	6	1.5	255	2	B82729	undecaprenyl pyrop
867	6	1.5	241	2	D87494	940	6	1.5	255	2	T50916	hypothetical prote
868	6	1.5	241	2	A70874	941	6	1.5	255	2	H64648	alternative transc
869	6	1.5	241	2	F95355	942	6	1.5	255	2	F71938	RNA polymerase sig
870	6	1.5	241	2	AH3543	943	6	1.5	255	2	A40613	avirulence protein
871	6	1.5	242	1	MF1VC1	944	6	1.5	255	2	C97081	2-Hydroxy-6-Oxo-6-
872	6	1.5	242	2	AC1542	945	6	1.5	255	2	H97394	orotidine 5' monop
873	6	1.5	242	2	D87346	946	6	1.5	255	2	AD1946	hypothetical prote
874	6	1.5	242	2	C83230	947	6	1.5	256	2	T31268	4-oxalocrotonate d
875	6	1.5	242	2	F84315	948	6	1.5	256	2	T35795	hypothetical prote
876	6	1.5	242	2	E70248	949	6	1.5	256	2	D83336	probable thiol-dis
877	6	1.5	242	2	A81782	950	6	1.5	256	2	G70666	hypothetical prote
878	6	1.5	242	2	C81205	951	6	1.5	256	2	F47031	bialaphos biosynth
879	6	1.5	243	2	E70846	952	6	1.5	256	2	T48588	hypothetical prote
880	6	1.5	243	2	B95855	953	6	1.5	257	2	AD2934	short chain denydr
881	6	1.5	243	2	H86487	954	6	1.5	257	2	AG2750	3-hydroxyacyl-CoA
882	6	1.5	243	2	AH2640	955	6	1.5	257	2	C98348	reductase (Al13282
883	6	1.5	244	2	B97423	956	6	1.5	257	2	F97531	probable short-cha
884	6	1.5	244	2	G95970	957	6	1.5	257	2	AH3584	cytline-binding pe
885	6	1.5	244	2	S48482	958	6	1.5	257	2	B86967	conserved hypotnet
886	6	1.5	245	2	AF1070	959	6	1.5	257	2	S16865	gene F41 protein -
887	6	1.5	245	2	H72612	960	6	1.5	257	2	B84346	hypothetical prote
888	6	1.5	245	2	H86536	961	6	1.5	257	2	B75510	hypothetical prote
889	6	1.5	245	2	G72086	962	6	1.5	257	2	C89945	conserved hypotnet
890	6	1.5	246	1	S30274	963	6	1.5	258	2	C84658	floral homeodomin
891	6	1.5	246	1	UX0230	964	6	1.5	258	2	T52377	p1B5 protein (imp
892	6	1.5	246	2	H95408	965	6	1.5	258	2	D95910	probable membrane
893	6	1.5	246	2	AB2062	966	6	1.5	258	2	JC2567	UDP-N-acetylmuramo
894	6	1.5	246	2	B69766	967	6	1.5	258	2	T25958	hypothetical prote
895	6	1.5	246	2	E39534	968	6	1.5	258	2	A10593	ABC transporter AT
896	6	1.5	247	1	A25852	969	6	1.5	259	2	T34637	probable type IV p
897	6	1.5	247	2	AD0757	970	6	1.5	259	2	B69113	cell division inhi
898	6	1.5	247	2	S60307	971	6	1.5	259	2	A13501	orotidine 5'-phosp
899	6	1.5	247	2	T52511	972	6	1.5	260	2	G87408	glucanate 5-denydr
900	6	1.5	247	2	G90348	973	6	1.5	260	2	D96004	probable 3-oxoacyl
901	6	1.5	248	1	PAOFCS	974	6	1.5	260	2	A10934	glutamate racemase
902	6	1.5	248	2	B87425	975	6	1.5	260	2	B83675	enoyl-CoA hydratase
903	6	1.5	248	2	A39534	976	6	1.5	260	2	B87392	conserved hypotnet
904	6	1.5	249	1	S19187	977	6	1.5	261	2	C81368	probable glutamine
905	6	1.5	249	2	AB2262	978	6	1.5	261	2	S72895	hypothetical prote

979 6 1.5 262 2 AC2644 flagellar basal-bo
980 6 1.5 262 2 C97426 flagellar basal-bo
981 6 1.5 262 2 B82950 chromosome partiti
982 6 1.5 262 2 A75009 probable aryl phos
983 6 1.5 262 2 C70587 hypothetical prote
984 6 1.5 262 2 T19323 hypothetical prote
985 6 1.5 262 2 B70699 hypothetical prote
986 6 1.5 263 1 SNRTC2 proteasome endopep
987 6 1.5 263 2 A83956 flagellar hook pro
988 6 1.5 263 2 H87296 aminomethyltransfe
989 6 1.5 263 2 C84598 probable ER lumen
990 6 1.5 264 2 B69535 DNA-directed RNA p
991 6 1.5 264 2 H83170 conserved hypothet
992 6 1.5 264 2 B70546 probable pnp prote
993 6 1.5 264 2 T34953 sporulation protei
994 6 1.5 264 2 G87356 hypothetical prote
995 6 1.5 264 2 AD3298 sensory transducti
996 6 1.5 265 1 JC5530 T-cluster binding
997 6 1.5 265 1 G69084 conserved hypothet
998 6 1.5 265 2 C82358 glutamate racemase
999 6 1.5 265 2 T07800 inositol-1(or 4)-m
1000 6 1.5 265 2 C84354 hypothetical prote

ALIGNMENTS

RESULT 1

membrane fusion protein NMA1970 [imported] - Neisseria meningitidis (strain Z2491 serogr
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81825
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H81825
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:G7380371; PIDN:CAB85190.1; PID:G738060
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: mtrC; NMA1970

Query Match 41.5%; Score 171; DB 2; Length 412;
Best Local Similarity 99.6%; Pred. No. 2.6e-159;
Matches 271; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFYAFKMRRAAALAAVALVYSSCGKGDAAOGGPPAREAPAPVGVVTHPQTVALT 60
DB 1 MAFYAFKMRRAAALAAVALVYSSCGKGDAAOGGPPAREAPAPVGVVTHPQTVALT 60
QY 61 VELPGLESLRTADYAAOVGGIIOKRLFOEGSVYRAGCPLOYIDSTVEANTESARAQLA 120
DB 61 VELPGLESLRTADYAAOVGGIIOKRLFOEGSVYRAGCPLOYIDSTVEANTESARAQLA 120
QY 121 TAOATLAKADADLARYKPLVAEAVEROEYDAAVTAKRSAAEGVKAQAIAKSAGINLNR 180
DB 121 TAOATLAKADADLARYKPLVAEAVEROEYDAAVTAKRSAAEGVKAQAIAKSAGINLNR 180
QY 181 SRTITPISGFIQSQKSEGTLLNAGDTTLATIRONPMYVWVWQSASVWMLRQIAAG 240
DB 181 SRTITPISGFIQSQKSEGTLLNAGDTTLATIRONPMYVWVWQSASVWMLRQIAAG 240
QY 241 KLLAADGVAVGAKFDGTVYPEKGRLLPADP 272
DB 241 KLLAADGVAVGAKFDGTVYPEKGRLLPADP 272

RESULT 2

F81051
membrane fusion protein NMB1716 [imported] - Neisseria meningitidis (strain MCS8 serogr
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: F81051
R:Teitelin, H.; Saunders, N.J.; Heideberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
rt, H.; Qin, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Piazza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.
A:Reference number: A81000; MUID:20175753; PMID:10710307
A:Accession: F81051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <TEP>
A:Cross-references: GB:AE002521; GB:AE002098; NID:G7226962; PIDN:AAF42063.1; PID:G722697
A:Experimental source: serogroup B, strain MCS8
C:Genetics:
A:Gene: NMB1716

Query Match 33.3%; Score 133; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 4.8e-122;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 VNESTGQITLRAAVPNDQNIIMPGLYVRVLMDOVAVDNAFVVPQOAVTRGAKDTVMIVNA 333
DB 274 VNESTGQITLRAAVPNDQNIIMPGLYVRVLMDOVAVDNAFVVPQOAVTRGAKDTVMIVNA 333
QY 334 OGMEPREVTVAAOOGTMYITVTSGLKODKDYVVRGISTAGITGAKKVTPKEMASSENOAA 393
DB 334 OGMEPREVTVAAOOGTMYITVTSGLKODKDYVVRGISTAGITGAKKVTPKEMASSENOAA 393
QY 394 APOGSGVTASEAK 406
DB 394 APOGSGVTASEAK 406

RESULT 3

S42418
probable antibiotic resistance protein mtrC - Neisseria gonorrhoeae (strain CH95) (fragm
C:Species: Neisseria gonorrhoeae
C:Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 08-Oct-1999
C:Accession: S42418; S40252
R:Pan, W.; Spratt, B.G.
Mol. Microbiol. 11, 769-775, 1994
A:Title: Regulation of the permeability of the gonococcal cell envelope by the mtr syste
A:Reference number: S42417; MUID:94254732; PMID:8196548
A:Accession: S42418
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-271 <PAR>
A:Cross-references: EMBL:Z25796; NID:G438190; PIDN:CAB1046.1; PID:G438192
A:Experimental source: CH95
C:Genetics:
A:Gene: mtrC
C:Superfamily: lipoyl/biotin-binding homology
F:68-111,183-211/Domains: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 24.3%; Score 100; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 7.7e-90;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 OEYDAAVTAKSAEAGVKAQAIAKSAGININRSRTITPISGFIQSQKSEGTLLNAGDT 207
DB 148 OEYDAAVTAKSAEAGVKAQAIAKSAGININRSRTITPISGFIQSQKSEGTLLNAGDT 207
QY 208 TVLATIRONPMYVWVWQSASVWMLRQIAEGKLLAADG 247
DB 208 TVLATIRONPMYVWVWQSASVWMLRQIAEGKLLAADG 247

RESULT 4

ES8757
 Probable efflux pump Z2509 [imported] - Escherichia coli (strain O157:H7, substrain EDL9)
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: ES8757
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: ES8757
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-373 <STO>
 A/Cross-references: GB:AE005174; NID:g12515494; PIDN:AA656521.1; GSPDB:GN00145; UWGP:Z25
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: Z2509

Query Match 2.7%; Score 11; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 78 QVGGIIQKRLF 88
 |||||
 Db 63 QVGGIIQKRLF 73

RESULT 5
 G90861
 Probable efflux pump Ecs1863 [similarity] - Escherichia coli (strain O157:H7, substrain
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
 C/Accession: G90861
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Sasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: G90861
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-373 <HAV>
 A/Cross-references: GB:BA000007; PIDN:BA835286.1; PID:g13361328; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RMD 0509952
 C/Genetics:
 A/Gene: Ecs1863

Query Match 2.7%; Score 11; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 78 QVGGIIQKRLF 88
 |||||
 Db 63 QVGGIIQKRLF 73

RESULT 6
 E87349
 H1YD family secretion protein [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: E87349
 R/Nierman, W.C.; Feldlym, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Land, M.T.; DeBoy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kolor
 n, J.; Brmlaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of Caulobacter crescentus.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: E87349
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-388 <STO>

A/Cross-references: GB:AE005673; NID:g13422055; PIDN:AAK22793.1; GSPDB:GN00148
 C/Genetics:
 A/Gene: CC0808

Query Match 2.7%; Score 11; DB 2; Length 388;
 Best Local Similarity 100.0%; Pred. No. 0.021;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 94 VRAGQPLYQID 104
 |||||
 Db 89 VRAGQPLYQID 99

RESULT 7
 AH0346
 putative H1YD family secretion protein [imported] - Yersinia pestis (strain CO92)
 C/Species: Yersinia pestis
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C/Accession: AH0346
 R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A/Reference number: AB0001; MUID:21470413; PMID:11586360.
 A/Accession: AH0346
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-444 <KTR>
 A/Cross-references: GB:AL590842; PIDN:CAC92099.1; PID:g15980817; GSPDB:GN00175
 C/Genetics:
 A/Gene: YPO2847

Query Match 2.2%; Score 9; DB 2; Length 444;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 181 SRIAPISG 189
 |||||
 Db 198 SRIAPISG 206

RESULT 8
 C81232
 50S ribosomal protein L24 NMB0153 [imported] - Neisseria meningitidis (strain MC58 serog
 C/Species: Neisseria meningitidis
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
 C/Accession: C81232; D82004
 R/Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, U.; Scarlato, V.; Maignani, V.; Pizze, M.
 Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
 A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: C81232
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-107 <TEP>
 A/Cross-references: GB:AE002373; GB:AE002098; NID:g7225359; PIDN:AAE40611.1; PID:g722537
 A/Experimental source: serogroup B, strain MC58
 R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churche, C.; Klee, S.R.; Morel
 Holroyd, S.; Jørgen, K.; Leach, S.; Moutle, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
 A/Reference number: A81775; MUID:20222556; PMID:10761919
 A/Accession: D82004
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-107 <PAR>
 A/Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83433.1; PID:g737889
 A/Experimental source: serogroup A, strain 22491
 C/Genetics:

A:Gene: rplX; NMB0153; NMA0118
C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 1.9%; Score 8; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 GDKVVEGI 368
DB 29 GDKVVEGI 36

RESULT 9

ribosomal protein L24 - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 04-Mar-1992 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999

C:Accession: E42645; C71506

R:Kaul, R.; Gray, G.J.; Koehncke, N.R.; Gu, L.J.

J. Bacteriol. 174, 1205-1212, 1992

A:Title: Cloning and sequence analysis of the Chlamydia trachomatis spc ribosomal protei

A:Reference number: A42645; MUID:92138612; PMID:1735714

A:Accession: E42645

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-111 <KAU>

A:Cross-references: GB:M80325; NID:9144617; PID:AAA23173.1; PID:9144622

A:Note: sequence extracted from NCBI backbone (NCBIN:79464; NCBI:79469)

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac

A:Reference number: A71570; MUID:9900809; PMID:9784136

A:Accession: C71506

A:Molecule type: DNA

A:Residues: 1-84,'R',86-104,'S',106-111 <ARN>

A:Cross-references: GB:AE00123; GB:AE001273; NID:G3328931; PID:AAE6118.1; PID:G332895

A:Experimental source: serotype D, strain UW-3/CX

C:Genetics:

A:Gene: r124

C:Superfamily: Escherichia coli ribosomal protein L24

C:Keywords: protein biosynthesis; ribosome

Query Match 1.9%; Score 8; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
DB 32 DKVVEGI 39

RESULT 10

ribosomal protein L24 TC0804 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000

C:Accession: D81664

R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: D81664

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-111 <TET>

A:Cross-references: GB:AE002347; GB:AE002160; NID:G7190815; PID:AAF39607.1; PID:G719083

A:Experimental source: strain Nigg (Mopn)

C:Genetics:

A:Gene: TC0804

C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 1.9%; Score 8; DB 2; Length 111;

Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
DB 32 DKVVEGI 39

RESULT 11

D70730

hypothetical protein Rv2272 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: D70730

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Rogers, J.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70730

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-122 <CDL>

A:Cross-references: GB:Z77163; GB:AL123456; NID:G3261610; PID:CA800963.1; PID:e255069;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv2272

Query Match 1.9%; Score 8; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LAAVAVLV 21
DB 37 LAAVAVLV 44

RESULT 12

3-oxoacyl-acyl-carrier protein reductase homolog yusR - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: C70022

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bexter

C.; Bron, S.; Brunnelle, S.; Bruschini, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Cho

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galler

lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinot,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauele

Y. M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanton,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpeira, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Wintere, P.; Wipat, A.; Yamamoto, H.; Yabane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: C70022

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-129 <KUN>

A:Cross-references: GB:Z99120; GB:AL009126; NID:G2635613; PID:CA815279.1; PID:G2635786

A:Experimental source: strain 168

C:Genetics:

A:Gene: yusR

C:Superfamily: ribitol dehydrogenase, short-chain alcohol dehydrogenase homology

F.1-71/Domain: short-chain alcohol dehydrogenase homology #status atypical <SABH>

Query Match 1.9%; Score 8; DB 2; Length 129;

Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 SIAGITGA 377
|||||||
Db 28 SIAGITGA 35

RESULT 13

D90703

probable fimbrial-like protein [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: D90703

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene A:Reference number: A59625; PMID:21156231; PMID:11258796

A:Accession: D90703

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-171 <HAV>

A:Cross-references: GB:BA000007; PIDN:BA034019.1; PID:g13360054; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: EC80596

C:Superfamily: type 1 fimbrial protein

Query Match 1.9%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 L1NAGDTT 208
|||||||
Db 61 L1NAGDTT 68

RESULT 14

G85553

probable fimbrial protein sfmF [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: G85553

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; PMID:21074935; PMID:11206551

A:Accession: G85553

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-171 <STO>

A:Cross-references: GB:AE005174; NID:g12513434; PIDN:AAG54891.1; GSPDB:GN00145; UWGP:Z06

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: sfmF

C:Superfamily: type 1 fimbrial protein

Query Match 1.9%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 L1NAGDTT 208
|||||||
Db 61 L1NAGDTT 68

RESULT 15

E64785

fimbrial protein homolog sfmF - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: E64785
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

A:Accession: E64785

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-171 <BLAT>

A:Cross-references: GB:AE000159; GB:U00096; NID:g1786739; PIDN:AACT73636.1; PID:g1786746;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: sfmF; ybcG

C:Superfamily: type 1 fimbrial protein

Query Match 1.9%; Score 8; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 201 L1NAGDTT 208
|||||||
Db 61 L1NAGDTT 68

RESULT 16

A70673

hypothetical protein Rv2980 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70673

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987; PMID:9634230

A:Accession: A70673

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-181 <COL>

A:Cross-references: GB:Z63018; GB:AL123456; NID:g3261671; PIDN:CAB05432.1; PID:e283361;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv2980

Query Match 1.9%; Score 8; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALAAV 18
|||||||
Db 16 AAALAAV 23

RESULT 17

A72499

hypothetical protein APE1967 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: A72499

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hakiawa, Y.; Jinno, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; PMID:9310339; PMID:10382966

A:Accession: A72499

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-183 <KAW>

A:Cross-references: DDBJ:AP000063; NID:g5105654; PIDN:BA080977.1; PID:g5105665

A:Experimental source: strain K1

C:Genetics:
A:Gene: APE1967
C:Superfamily: conserved hypothetical protein MTH986

Query Match 1.9%; Score 8; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVAL 20
DB 118 ALAAVAL 125

RESULT 18

G90849
hypothetical protein Ecs1767 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90849
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A59629; MUID:21156231; PMID:11258796
A:Accession: G90849
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA035190.1; PID:g13361232; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:
A:Gene: Ecs1767

Query Match 1.9%; Score 8; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LIFADPVV 274
DB 130 LIFADPVV 137

RESULT 19

R85707
unknown protein encoded by prophage CP-9330 [imported] - Escherichia coli (strain O157:H
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85707
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimmlanta, E.; Potamouotis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85707
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <STO>
A:Cross-references: GB:AB005174; NID:g12514992; PIDN:AA056122.1; GSPDB:GN00145; UMGF:Z20
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2047

Query Match 1.9%; Score 8; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LIFADPVV 274
DB 130 LIFADPVV 137

RESULT 20

T36757

probable secreted lipase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36757
R:Saunders, D.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21613
A:Accession: T36757
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-290 <SAU>
A:Cross-references: EMBL:AL096849; PIDN:CAB50950.1; GSPDB:GN00070; SCODDB:SC11.24c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODDB:SC11.24c

Query Match 1.9%; Score 8; DB 2; Length 290;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LAAVALV 21
DB 16 LAAVALV 23

RESULT 21

C95961
probable sugar uptake ABC transporter periplasmic solute-binding protein precursor SMD21
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: C95961
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,663-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: C95961
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <KUB>
A:Cross-references: GB:AU591985; PIDN:CNC49355.1; PID:g15140841; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeJaune,
hebaull, P.; Vandenhof, W.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD21377
A:Genome: plasmid

Query Match 1.9%; Score 8; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAVA 19
DB 216 AALAAVA 223

RESULT 22

H86740
transcription regulator yJfE [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 06-Jan-2003
C:Accession: H86740
R:Boilotin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: H86740
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <STO>
A:Cross-references: GB:AE005176; PID:g12723863; PID:AAK05026.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yjfe
C:Superfamily: Bacillus subtilis probable transcription regulator yjhJ

Query Match 1.9%; Score 8; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 TLNAGDT 207
|||
Db 202 TLNAGDT 209

RESULT 23
C70415
cation efflux system (czcB-like) - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: C70415
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O.V.
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98198666; PMID:9537320
A:Accession: C70415
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-348 <AOE>
A:Cross-references: GB:AE000735; NID:g2983749; PID:AA07317.1; PID:g2983757; GB:AE00065
A:Experimental source: strain V85
C:Genetics:
A:Gene: czcB2

Query Match 1.9%; Score 8; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 LKGDGVV 365
|||
Db 325 LKGDGVV 332

RESULT 24
AC0961
probable ATP/GTP-binding protein STY3967 [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0961
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Moule, S.; O'Gaora, P.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi.
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0961
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <PAR>
A:Cross-references: GB:AL513382; PID:CA03183.1; PID:G16504816; GSPDB:GN00176
C:Genetics:
A:Gene: STY3967
C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match 1.9%; Score 8; DB 2; Length 408;
Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 VGVVTVHP 54
|||
Db 64 VGVVTVHP 71

RESULT 25
B65171
hypothetical protein yidR - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: B65171
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co.A.; Rose, D.U.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B65171
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-416 <BLAT>
A:Cross-references: GB:AE000446; GB:U00096; NID:g2367261; PID:AA076712.1; PID:g1790124;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yidR
C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match 1.9%; Score 8; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 VGVVTVHP 54
|||
Db 76 VGVVTVHP 83

RESULT 26
E91207
hypothetical protein ECG4629 [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: E91207
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno.
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <HAY>
A:Cross-references: GB:BA000007; PID:BA038052.1; PID:g13364104; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECG4629
C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match 1.9%; Score 8; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 VGVVTVHP 54
|||
Db 76 VGVVTVHP 83

RESULT 27
G86053
hypothetical protein yidR [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G86053

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Polanousis, K.; Apodaca,
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G86053
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-416 <STO>

A:Cross-references: GB:AE005174; NID:912518530; PIDN:AAG58891.1; GSPDB:GN00145; UMGF:251
 A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:
 A:Gene: yldR

C:Superfamily: *Escherichia coli* hypothetical protein yldR

Query Match 1.9%; Score 8; DB 2; Length 416;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVVTVHP 54
 |||||
 DB 76 VGVVTVHP 83

RESULT 28

MMBER3
 UA43 Protein - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1
 C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000

C:Accession: G30088
 R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; DoJan, A.; Frame, M.C.; McNab, D.; Perz

J. Gen. Virol. 69, 1531-1574, 1988

A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim
 A:Reference number: A30083; MUID:88274327; PMID:2839594

A:Accession: G30088
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-434 <MCG>

A:Cross-references: GB:X14112; NID:G1944536; PIDN:CAA32306.1; PID:G59543; GB:D00317
 C:Genetics:

A:Gene: UL43
 C:Superfamily: herpesvirus UL43 protein

Query Match 1.9%; Score 8; DB 1; Length 434;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAATAAAV 18
 |||||
 DB 199 AAATAAAV 206

RESULT 29

DDHUM2
 methionyl aminopeptidase (EC 3.4.11.18) 2 - human

N:Alternate names: p67; translation initiation factor eIF-2-associated protein
 C:Species: Homo sapiens (man)

C:Date: 14-Jul-1995 #sequence_revision 23-Aug-1996 #text_change 01-Feb-2002
 C:Accession: S52112

R:Li, X.; Chang, Y.H.
 Biochim. Biophys. Acta 1260, 333-336, 1995

A:Title: Molecular cloning of a human complementary DNA encoding an initiation factor 2-
 A:Reference number: S52112; MUID:95178556; PMID:7873610

A:Accession: S52112
 A:Molecule type: mRNA

A:Residues: 1-478
 A:Cross-references: GB:U13261; NID:G687242; PIDN:AAC63402.1; PID:G687243

C:Genetics:
 A:Gene: GDB:P67EIF2

A:Cross-references: GDB:512821
 C:Function:

A:Description: catalyzes hydrolysis of amino-terminal methionine from proteins
 C:Superfamily: human methionyl aminopeptidase

C:Keywords: aminopeptidase; cobalt; metalloprotein; protein biosynthesis
 F:251,262,459/Binding site: cobalt 2 (Asp, Asp, Glu) #status predicted
 F:262,331,364,459/Binding site: cobalt 1 (Asp, His, Glu, Glu) #status predicted

Query Match 1.9%; Score 8; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 30

A46702
 methionyl aminopeptidase (EC 3.4.11.18) 2 - rat

N:Alternate names: p67; translation initiation factor eIF2-associated protein
 C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 01-Feb-2002
 C:Accession: A46702

R:Mu, S.; Gupta, S.; Chatterjee, N.; Hilleman, R.E.; Kinzy, T.G.; Denalow, N.D.; Merrick,
 J. Biol. Chem. 268, 10796-10801, 1993

A:Title: Cloning and characterization of complementary DNA encoding the eukaryotic initi
 A:Reference number: A46702; MUID:93266517; PMID:8496145

A:Accession: A46702
 A:Molecule type: mRNA

A:Residues: 1-480 <WUA>
 A:Cross-references: GB:L10652; NID:G204003; PIDN:AAA41111.1; PID:G204004

C:Superfamily: human methionyl aminopeptidase
 C:Keywords: aminopeptidase

Query Match 1.9%; Score 8; DB 2; Length 480;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 31

T22088
 hypothetical protein F42D1.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T22088
 R:Lightning, J.

submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19511

A:Accession: T22088
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-489 <WII>

A:Cross-references: EMBL:Z81081; PIDN:CAB03091.1; GSPDB:GN00028; CESP:F42D1.3
 A:Experimental source: clone F42D1

C:Genetics:
 A:Gene: CESP:F42D1.3

A:Map position: X
 A:Intons: 21/3; 101/3; 159/2; 201/2; 228/3; 261/2; 322/2; 350/3; 393/3; 420/2

Query Match 1.9%; Score 8; DB 2; Length 489;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALVLS 23
 |||||
 DB 115 AAVALVLS 122

RESULT 32

AB1070
 probable carbon starvation protein [imported] - *Salmonella enterica* subsp. *enterica* sero

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi

A>Note: this species has also been called *Salmonella typhi*.
 C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AB1070
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Cronin, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Author: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AB1070
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-716 <PAR>
 A:Cross-References: GB:AL513382; PIDN:CAD03377.1; PID:g16505647; GSPDB:GN00176
 C:Genetics: cstA
 A:Gene: cstA
 C:Superfamily: carbon starvation protein

Query Match 1.9%; Score 8; DB 2; Length 716;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AAVALVS 23
 |||||
 Db 577 AAVALVS 584

RESULT 33

probable integral membrane export protein - *Streptomyces coelicolor*
 C:Species: *Streptomyces coelicolor*
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35818
 R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, February 1999
 A:Reference number: Z21589
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-748 <MUR>
 A:Cross-References: EMBL:AL035569; PIDN:CAB37578.1; GSPDB:GN00070; SCOEDB:SCB9.14
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCB9.14

Query Match 1.9%; Score 8; DB 2; Length 748;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ALAAVAL 20
 |||||
 Db 666 ALAAVAL 673

RESULT 34

membrane alanyl aminopeptidase (EC 3.4.11.2) [imported] - *Yersinia pestis* (strain CO92)
 AH0172
 C:Species: *Yersinia pestis*
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 01-Feb-2002
 C:Accession: AH0172
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AH0172
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-871 <KUR>
 A:Cross-References: GB:AL590842; PIDN:CAC90243.1; PID:g15979463; GSPDB:GN00175
 C:Genetics:

A:Gene: pepN
 A:Superfamily: microsomal aminopeptidase
 C:Keywords: aminopeptidase

Query Match 1.9%; Score 8; DB 2; Length 871;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ALAAAVA 19
 |||||
 Db 715 ALAAAVA 722

RESULT 35

DEEC
 acetaldehyde dehydrogenase (acetylating) (EC 1.2.1.10) / alcohol dehydrogenase (EC 1.1.1 N/Alternate names: acetaldehyde/alcohol dehydrogenase; Adh; aldehyde reductase
 N/Contains: pyruvate-formate-lyase deactivase
 C:Species: *Escherichia coli*
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 01-Mar-2002
 C:Accession: J50406; S14809; I76913; D64871; S23871
 R:Goodlove, P.E.; Cunningham, P.R.; Parker, J.; Clark, D.P.
 Gene 85, 209-214, 1989
 A:Title: Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding A:Reference number: J50406; MUID:90152365; PMID:2693398
 A:Accession: J50406
 A:Molecule type: DNA
 A:Residues: 1-891 <GOO>
 A:Cross-References: GB:M3504; NID:g415205; PIDN:AAA23420.1; PID:g145206
 A:Experimental source: plasmid PH18
 A>Note: residues 2-11 were confirmed by protein sequencing
 R:Kessler, D.; Leibrecht, I.; Knappe, J.
 FEBS Lett. 281, 59-63, 1991

A:Title: Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escher A:Reference number: S14809; MUID:91200315; PMID:2015910*
 A:Accession: S14809
 A:Molecule type: DNA
 A:Residues: 1-891 <KES>
 A:Cross-References: EMBL:X59263; NID:g40899; PIDN:CAA1955.1; PID:g40900
 A>Note: only nucleotide sequences flanking the coding region are shown
 A:Note: enzyme assays demonstrated stimulation by ferrous ion when either acetyl-CoA or R:Danchn, A.; Kryn, E.
 Microbiology 141, 959-960, 1995
 A:Title: Filling the gap between hns and adh in *Escherichia coli* K12.
 A:Reference number: I57117; MUID:95291445; PMID:7773397
 A:Accession: I76913
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 849-891 <DAN>
 A:Cross-References: EMBL:X67326; NID:g43077; PIDN:CAA47743.1; PID:g43081
 A>Note: submitted to the EMBL Data Library, July 1992
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co A.; Rose, D.D.; Mau, B.; Sha, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D64871
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-891 <BAT>
 A:Cross-References: GB:AE000222; GB:U00096; NID:g1787486; PIDN:AACT4323.1; PID:g1787493;
 A:Experimental source: strain K-12, substrain M01655
 C:Genetics:
 A:Gene: adhE
 C:Complex: homomultimer containing more than 40 chains
 C:Function: <ADH>
 A:Description: catalyzes the reduction of acetyl-CoA to enzyme-bound thiohemiacetal usin A:Pathway: anaerobic glucose fermentation
 A>Note: operates only in the absence of nitrate
 C:Function: <ARD>
 A:Description: catalyzes the reduction of the thiohemiacetal to ethanol using NADH A:Pathway: anaerobic glucose fermentation
 A>Note: operates only in the absence of nitrate; generates ethanol and NAD+

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact
C:Keywords: alcohol metabolism; coenzyme A, homomultimer; iron; multifunctional enzyme;
F:2-891/Product: acetaldehyde/alcohol dehydrogenase #status predicted <MAT>
F:8-266/Domain: aldehyde dehydrogenase homology #link AADH <ALDH>
F:454-858/Domain: lactaldehyde reductase homology #link ARD <LAR>
F:246/Active site: Cys #link AADH #status predicted

Query Match 1.9%; Score 8; DB 1; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||
DB 36 RAALALAA 43

RESULT 36

DB5704
hypothetical protein adhe [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: DB5704

R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimmlanta, E.; Potamoustis, K.; Apodaca,
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; PMID:21074935; PMID:11206551

A:Accession: DB5704

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <STO>

A:Cross-references: GB:AE005174; NID:G12514960; PIDN:AG56096.1; GSPDB:GN00145; UMG:Z20

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics: adhe

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

Query Match 1.9%; Score 8; DB 2; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||
DB 36 RAALALAA 43

RESULT 37

DB50846
acetaldehyde dehydrogenase (acetylating) (EC 1.2.1.10) / alcohol dehydrogenase (EC 1.1.1

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 27-Nov-2001

C:Accession: E90846

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gisawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; PMID:21156231; PMID:11258796

A:Accession: E90846

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA03164.1; PID:G13361206; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECA1741

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

C:Keywords: coenzyme A; oxidoreductase

Query Match 1.9%; Score 8; DB 2; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17

DB 36 RAALALAA 43
|||||

RESULT 38

AC0265
alcohol dehydrogenase (EC 1.1.1.1) [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 11-Jan-2002

C:Accession: AC0265

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlhal, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; PMID:21470413; PMID:11586360

A:Accession: AC0265

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC90987.1; PID:G15980182; GSPDB:GN00175

C:Genetics:

A:Gene: adhe

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

C:Keywords: oxidoreductase

Query Match 1.9%; Score 8; DB 2; Length 891;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||
DB 36 RAALALAA 43

RESULT 39

AE0650
alcohol dehydrogenase (EC 1.1.1.1) [imported] - Salmonella enterica subsp. enterica sero

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AE0650

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero

A:Reference number: AB0502; PMID:21534947; PMID:11677608

A:Accession: AE0650

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-892 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08384.1; PID:G16502428; GSPDB:GN00176

C:Genetics:

A:Gene: adh

C:Superfamily: acetaldehyde/alcohol dehydrogenase; aldehyde dehydrogenase homology; lact

C:Keywords: oxidoreductase

Query Match 1.9%; Score 8; DB 2; Length 892;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
|||||
DB 36 RAALALAA 43

RESULT 40

T01566

hypothetical protein A_T01566.23 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 04-Mar-2000

C/Accession: T01566
R/Dempsey, S.; Harper, M.
Submitted to the EMBL Data Library, July 1997
A/Description: The sequence of A. thaliana TW018A10.
A/Reference number: Z14348
A/Accession: T01566
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-979 <DEM>
A/Cross-references: EMBL:AF013294; NID:g2252848; PID:g2252867
A/Experimental source: cultivar Columbia
C/Genetics:
A/Map position: 4
A/Introns: 466/3; 569/3; 649/3; 688/1; 740/3; 877/3
A/Note: A TW018A10.23
C/Superfamily: Arabidopsis thaliana hypothetical protein A_TW018A10.23

Query Match 1.9%; Score 8; DB 2; Length 979;
Best Local Similarity 100.0%; Pred.No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 AALAAVA 19
Db 117 AALAAVA 124

Search completed: September 8, 2003, 14:09:40
Job time : 37 secs

THIS PAGE BLANK (USPTO

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:07:27 ; Search time 21 Seconds
(without alignments)
830.099 Million cell updates/sec

Title: US-09-889-756a-2

Perfect score: 412

Sequence: 1 MAFYAFKAMRAAALAAVAAL.....AAPQSGVOTASEAKTASBAE 412

Scoring table: OLIGO

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: Issued Patents, AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.4	124	4	US-09-461-325-398
2	9	2.2	447	4	US-09-328-352-7383
3	8	1.9	111	4	US-09-732-210-240
4	8	1.9	111	4	US-09-732-210-764
5	8	1.9	194	4	US-09-461-325-397
6	8	1.9	343	4	US-09-252-991A-24458
7	8	1.9	478	2	US-09-040-799-3
8	8	1.9	478	2	US-09-093-448-1
9	8	1.9	478	3	US-09-093-448-2
10	8	1.9	478	3	US-09-093-448-3
11	8	1.9	478	4	US-09-813-555-1
12	8	1.9	478	4	US-09-813-555-2
13	8	1.9	478	4	US-09-813-555-3
14	15	1.7	19	3	US-09-010-999-11
15	14	1.7	24	4	US-09-227-357-347
16	17	1.7	111	4	US-09-198-452A-675
17	17	1.7	126	1	US-08-312-387B-2
18	17	1.7	126	1	US-08-683-452-2
19	17	1.7	126	1	US-08-683-452-2
20	17	1.7	126	2	US-08-878-360-2
21	17	1.7	126	3	US-08-478-140B-2
22	17	1.7	126	4	US-09-333-412-2
23	17	1.7	126	4	US-09-338-943-2
24	7	1.7	131	1	US-08-664-596B-18
25	7	1.7	132	1	US-08-534-975-4
26	7	1.7	132	2	US-08-954-470-4
27	7	1.7	132	3	US-09-129-855A-4

28	7	1.7	132	3	US-09-247-154-4	Sequence 4, Appli
29	7	1.7	132	4	US-09-480-718-4	Sequence 4, Appli
30	7	1.7	132	4	US-09-610-833-4	Sequence 4, Appli
31	7	1.7	132	4	US-09-129-855A-4	Sequence 4, Appli
32	7	1.7	140	4	US-09-252-991A-18327	Sequence 18327, A
33	7	1.7	169	1	US-08-145-995A-7	Sequence 7, Appli
34	7	1.7	169	2	US-08-451-747-7	Sequence 7, Appli
35	7	1.7	169	3	US-09-134-852-7	Sequence 7, Appli
36	7	1.7	195	4	US-09-328-352-6419	Sequence 6419, Ap
37	7	1.7	200	4	US-09-252-991A-25479	Sequence 25479, Ap
38	7	1.7	262	4	US-09-252-991A-22833	Sequence 22833, A
39	7	1.7	275	4	US-09-107-532A-4013	Sequence 4013, Ap
40	7	1.7	287	4	US-09-107-532A-5035	Sequence 5035, Ap
41	7	1.7	310	4	US-09-134-001C-5326	Sequence 5326, Ap
42	7	1.7	319	4	US-09-252-991A-32369	Sequence 32369, A
43	7	1.7	330	4	US-09-252-991A-18946	Sequence 18946, A
44	7	1.7	357	4	US-09-252-991A-24119	Sequence 24119, A
45	7	1.7	374	1	US-08-095-726-14	Sequence 14, Appli
46	7	1.7	374	4	US-08-096-623A-14	Sequence 32966, A
47	7	1.7	377	4	US-09-252-991A-32966	Sequence 4138, Ap
48	7	1.7	402	4	US-09-134-001C-4138	Sequence 4953, Ap
49	7	1.7	402	4	US-09-328-352-4953	Sequence 31, Appli
50	7	1.7	409	3	US-09-066-046-31	Sequence 19, Appli
51	7	1.7	409	4	US-09-066-047-19	Sequence 6858, Ap
52	7	1.7	409	4	US-09-328-352-6858	Sequence 30606, A
53	7	1.7	410	4	US-09-252-991A-30606	Sequence 16, Appli
54	7	1.7	414	4	US-09-252-991A-27975	Sequence 21608, A
55	7	1.7	418	1	US-08-615-170-16	Sequence 22487, A
56	7	1.7	422	4	US-09-252-991A-21608	Sequence 30927, A
57	7	1.7	440	4	US-09-252-991A-22487	Sequence 23608, A
58	7	1.7	458	4	US-09-252-991A-30337	Sequence 4, Appli
59	7	1.7	479	4	US-09-252-991A-23608	Sequence 4, Appli
60	7	1.7	489	1	US-08-434-702-4	Sequence 4, Appli
61	7	1.7	492	3	US-09-006-632-4	Sequence 4, Appli
62	7	1.7	492	3	US-09-325-274-8	Sequence 8, Appli
63	7	1.7	492	4	US-09-006-632-8	Sequence 8, Appli
64	7	1.7	494	3	US-09-006-632-8	Sequence 8, Appli
65	7	1.7	494	3	US-09-325-274-8	Sequence 8, Appli
66	7	1.7	494	3	US-09-325-274-8	Sequence 8, Appli
67	7	1.7	496	3	US-09-348-443-2	Sequence 2, Appli
68	7	1.7	534	4	US-09-252-991A-30902	Sequence 30902, A
69	7	1.7	538	4	US-09-252-991A-22200	Sequence 22200, A
70	7	1.7	542	4	US-09-252-991A-31968	Sequence 31968, A
71	7	1.7	544	2	US-08-933-376A-4	Sequence 4, Appli
72	7	1.7	561	4	US-09-252-991A-23080	Sequence 23080, A
73	7	1.7	561	4	US-09-252-991A-31034	Sequence 31034, A
74	7	1.7	572	2	US-08-933-376A-2	Sequence 2, Appli
75	7	1.7	574	4	US-09-252-991A-18912	Sequence 18912, A
76	7	1.7	574	4	US-09-252-991A-31651	Sequence 31651, A
77	7	1.7	591	4	US-09-519-232-66	Sequence 66, Appli
78	7	1.7	597	4	US-09-252-991A-23152	Sequence 23152, A
79	7	1.7	610	4	US-09-252-991A-24992	Sequence 24942, A
80	7	1.7	615	4	US-09-996-243-194	Sequence 19, Appli
81	7	1.7	631	4	US-09-311-626B-16	Sequence 16, Appli
82	7	1.7	674	4	US-09-252-991A-24934	Sequence 24934, A
83	7	1.7	687	4	US-09-252-991A-26187	Sequence 26187, A
84	7	1.7	697	4	US-09-328-352-6212	Sequence 6212, Ap
85	7	1.7	705	4	US-09-252-991A-27442	Sequence 27442, A
86	7	1.7	711	4	US-09-252-991A-22259	Sequence 22259, A
87	7	1.7	738	3	US-08-864-038A-3	Sequence 3, Appli
88	7	1.7	742	4	US-09-252-991A-29239	Sequence 29239, A
89	7	1.7	757	4	US-09-252-991A-28180	Sequence 28180, A
90	7	1.7	801	4	US-09-252-991A-18228	Sequence 18228, A
91	7	1.7	1016	4	US-09-252-991A-18914	Sequence 18914, A
92	7	1.7	1057	3	US-08-931-820-4	Sequence 4, Appli
93	7	1.7	1078	3	US-08-963-825-21	Sequence 21, Appli
94	7	1.7	1078	4	US-09-500-811-21	Sequence 21, Appli
95	7	1.7	1078	4	US-09-570-573-21	Sequence 21, Appli
96	7	1.7	1078	4	US-09-548-608-21	Sequence 21, Appli
97	7	1.7	1144	3	US-08-726-214-6	Sequence 6, Appli
98	7	1.7	1186	1	US-08-485-568A-4	Sequence 4, Appli
99	7	1.7	1186	1	US-08-357-698-6	Sequence 6, Appli
100	7	1.7	1186	2	US-08-590-554A-4	Sequence 4, Appli

101	7	1.7	1186	2	US-09-184-223-4	Sequence 4, Appl	174	6	1.5	35	3	US-08-360-107A-49	Sequence 49, Appl
102	7	1.7	1186	5	PCT-US93-12682-6	Sequence 6, Appl	175	6	1.5	35	3	US-08-484-223B-19	Sequence 49, Appl
103	7	1.7	1206	4	US-09-252-991A-19632	Sequence 19632, A	176	6	1.5	35	3	US-08-484-223B-19	Sequence 147, Appl
104	7	1.7	1212	4	US-09-268-866-2	Sequence 2, Appl	177	6	1.5	35	3	US-08-919-597-49	Sequence 49, Appl
105	7	1.7	1700	4	US-09-252-991A-21763	Sequence 21763, A	178	6	1.5	35	3	US-08-919-597-147	Sequence 49, Appl
106	7	1.7	2284	4	US-09-252-991A-23547	Sequence 23547, A	179	6	1.5	35	3	US-08-475-668A-49	Sequence 49, Appl
107	7	1.7	2291	4	US-09-252-991A-21854	Sequence 21854, A	180	6	1.5	35	3	US-08-475-668A-147	Sequence 147, Appl
108	7	1.7	3079	5	PCT-US94-00198-4	Sequence 4, Appl	181	6	1.5	35	3	US-08-485-551A-49	Sequence 49, Appl
109	6	1.5	11	3	US-08-874-347-11	Sequence 11, Appl	182	6	1.5	35	3	US-08-485-551A-147	Sequence 147, Appl
110	6	1.5	11	2	US-09-093-522-11	Sequence 11, Appl	183	6	1.5	35	3	US-08-471-913A-147	Sequence 147, Appl
111	6	1.5	12	2	US-08-874-347-17	Sequence 17, Appl	184	6	1.5	35	3	US-08-485-546A-49	Sequence 49, Appl
112	6	1.5	13	5	US-09-093-522-17	Sequence 17, Appl	185	6	1.5	35	3	US-08-485-546A-147	Sequence 147, Appl
113	6	1.5	13	5	PCT-US95-04121-14	Sequence 14, Appl	186	6	1.5	35	3	US-09-082-279B-147	Sequence 147, Appl
114	6	1.5	14	1	US-08-232-453A-26	Sequence 26, Appl	187	6	1.5	35	3	US-09-082-279B-153	Sequence 153, Appl
115	6	1.5	16	1	US-08-346-849-21	Sequence 21, Appl	188	6	1.5	35	3	US-09-082-279B-155	Sequence 155, Appl
116	6	1.5	16	2	US-08-293-284A-21	Sequence 21, Appl	189	6	1.5	35	3	US-09-082-279B-155	Sequence 155, Appl
117	6	1.5	16	2	US-08-898-300-21	Sequence 21, Appl	190	6	1.5	35	3	US-09-082-279B-156	Sequence 156, Appl
118	6	1.5	16	3	US-08-504-538A-2	Sequence 2, Appl	191	6	1.5	35	3	US-09-082-279B-157	Sequence 157, Appl
119	6	1.5	20	3	US-09-249-458A-2	Sequence 2, Appl	192	6	1.5	35	3	US-09-082-279B-158	Sequence 158, Appl
120	6	1.5	20	4	US-08-630-052-2	Sequence 2, Appl	193	6	1.5	35	3	US-09-082-279B-159	Sequence 159, Appl
121	6	1.5	20	5	PCT-US95-09307-2	Sequence 2, Appl	194	6	1.5	35	3	US-09-082-279B-160	Sequence 160, Appl
122	6	1.5	23	2	US-08-455-968B-21	Sequence 21, Appl	195	6	1.5	35	3	US-09-082-279B-161	Sequence 161, Appl
123	6	1.5	24	3	US-08-504-538A-7	Sequence 7, Appl	196	6	1.5	35	3	US-09-082-279B-162	Sequence 162, Appl
124	6	1.5	24	3	US-09-249-458A-7	Sequence 7, Appl	197	6	1.5	35	3	US-09-082-279B-163	Sequence 163, Appl
125	6	1.5	24	4	US-08-630-052-7	Sequence 7, Appl	198	6	1.5	35	3	US-09-082-279B-164	Sequence 164, Appl
126	6	1.5	24	5	PCT-US95-09307-7	Sequence 7, Appl	199	6	1.5	35	4	US-08-474-349A-288	Sequence 288, Appl
127	6	1.5	25	2	US-08-934-741A-23	Sequence 23, Appl	200	6	1.5	35	4	US-08-474-349A-289	Sequence 289, Appl
128	6	1.5	25	3	US-09-049-691-58	Sequence 58, Appl	201	6	1.5	35	4	US-08-474-349A-147	Sequence 147, Appl
129	6	1.5	26	3	US-09-208-684-2	Sequence 8, Appl	202	6	1.5	35	4	US-08-474-349A-285	Sequence 285, Appl
130	6	1.5	27	3	US-08-809-397-9	Sequence 9, Appl	203	6	1.5	35	4	US-08-474-349A-286	Sequence 286, Appl
131	6	1.5	27	4	US-09-425-597-9	Sequence 9, Appl	204	6	1.5	35	4	US-08-474-349A-287	Sequence 287, Appl
132	6	1.5	27	5	PCT-US95-12502-9	Sequence 9, Appl	205	6	1.5	35	4	US-08-474-349A-288	Sequence 288, Appl
133	6	1.5	28	3	US-08-486-099-146	Sequence 146, App	206	6	1.5	35	4	US-08-474-349A-290	Sequence 290, Appl
134	6	1.5	28	3	US-08-484-223B-146	Sequence 146, App	207	6	1.5	35	4	US-08-474-349A-291	Sequence 291, Appl
135	6	1.5	28	3	US-08-919-597-146	Sequence 146, App	208	6	1.5	35	4	US-08-474-349A-292	Sequence 292, Appl
136	6	1.5	28	3	US-08-475-668A-146	Sequence 146, App	209	6	1.5	35	4	US-08-474-349A-293	Sequence 293, Appl
137	6	1.5	28	3	US-08-485-551A-146	Sequence 146, App	210	6	1.5	35	4	US-08-474-349A-294	Sequence 294, Appl
138	6	1.5	28	3	US-08-471-913A-146	Sequence 146, App	211	6	1.5	35	4	US-08-474-349A-295	Sequence 295, Appl
139	6	1.5	28	3	US-08-485-264A-146	Sequence 146, App	212	6	1.5	35	4	US-08-474-349A-296	Sequence 296, Appl
140	6	1.5	28	3	US-09-082-279B-36	Sequence 36, Appl	213	6	1.5	35	4	US-09-315-304B-34	Sequence 34, Appl
141	6	1.5	28	3	US-09-082-279B-37	Sequence 37, Appl	214	6	1.5	35	4	US-09-315-304B-153	Sequence 153, Appl
142	6	1.5	28	3	US-09-082-279B-38	Sequence 38, Appl	215	6	1.5	35	4	US-09-315-304B-154	Sequence 154, Appl
143	6	1.5	28	4	US-08-474-349A-146	Sequence 146, App	216	6	1.5	35	4	US-09-315-304B-155	Sequence 155, Appl
144	6	1.5	28	4	US-08-474-349A-318	Sequence 318, App	217	6	1.5	35	4	US-09-315-304B-156	Sequence 156, Appl
145	6	1.5	28	4	US-09-315-304B-36	Sequence 36, Appl	218	6	1.5	35	4	US-09-315-304B-157	Sequence 157, Appl
146	6	1.5	28	4	US-09-315-304B-37	Sequence 37, Appl	219	6	1.5	35	4	US-09-315-304B-162	Sequence 162, Appl
147	6	1.5	28	4	US-09-315-304B-38	Sequence 38, Appl	220	6	1.5	35	4	US-09-315-304B-163	Sequence 163, Appl
148	6	1.5	28	4	US-08-470-896-146	Sequence 146, App	221	6	1.5	35	4	US-09-315-304B-164	Sequence 164, Appl
149	6	1.5	28	4	US-08-485-546A-146	Sequence 146, App	222	6	1.5	35	4	US-09-315-304B-161	Sequence 161, App
150	6	1.5	28	4	US-09-834-784-36	Sequence 36, Appl	223	6	1.5	35	4	US-09-315-304B-162	Sequence 162, App
151	6	1.5	28	4	US-09-834-784-37	Sequence 37, Appl	224	6	1.5	35	4	US-09-315-304B-163	Sequence 163, App
152	6	1.5	28	4	US-09-834-784-38	Sequence 38, Appl	225	6	1.5	35	4	US-09-315-304B-164	Sequence 164, App
153	6	1.5	29	1	US-07-746-705A-2	Sequence 2, Appl	226	6	1.5	35	4	US-09-315-304B-165	Sequence 165, App
154	6	1.5	29	2	US-08-380-182-1	Sequence 1, Appl	227	6	1.5	35	4	US-08-470-896-49	Sequence 49, Appl
155	6	1.5	29	2	US-09-051-934-53	Sequence 53, Appl	228	6	1.5	35	4	US-08-485-546A-147	Sequence 147, Appl
156	6	1.5	30	3	US-09-051-934-54	Sequence 54, Appl	229	6	1.5	35	4	US-08-485-546A-147	Sequence 147, Appl
157	6	1.5	30	6	5188961-8	Patent No. 5188961	230	6	1.5	35	4	US-09-834-784-34	Sequence 34, Appl
158	6	1.5	34	2	US-08-934-741A-3	Sequence 3, Appl	231	6	1.5	35	4	US-09-834-784-153	Sequence 153, Appl
159	6	1.5	34	3	US-08-486-099-19	Sequence 19, Appl	232	6	1.5	35	4	US-09-834-784-154	Sequence 154, Appl
160	6	1.5	34	3	US-08-360-107A-19	Sequence 19, Appl	233	6	1.5	35	4	US-09-834-784-155	Sequence 155, Appl
161	6	1.5	34	3	US-08-484-223B-19	Sequence 19, Appl	234	6	1.5	35	4	US-09-834-784-156	Sequence 156, Appl
162	6	1.5	34	3	US-08-919-597-19	Sequence 19, Appl	235	6	1.5	35	4	US-09-834-784-157	Sequence 157, Appl
163	6	1.5	34	3	US-08-475-668A-19	Sequence 19, Appl	236	6	1.5	35	4	US-09-834-784-158	Sequence 158, Appl
164	6	1.5	34	3	US-08-485-551A-19	Sequence 19, Appl	237	6	1.5	35	4	US-09-834-784-159	Sequence 159, Appl
165	6	1.5	34	3	US-08-471-913A-19	Sequence 19, Appl	238	6	1.5	35	4	US-09-834-784-160	Sequence 160, Appl
166	6	1.5	34	3	US-08-485-264A-19	Sequence 19, Appl	239	6	1.5	35	4	US-09-834-784-161	Sequence 161, Appl
167	6	1.5	34	4	US-08-474-349A-19	Sequence 19, Appl	240	6	1.5	35	4	US-09-834-784-162	Sequence 162, Appl
168	6	1.5	34	4	US-08-470-896-19	Sequence 19, Appl	241	6	1.5	35	4	US-09-834-784-163	Sequence 163, Appl
169	6	1.5	34	4	US-08-485-546A-19	Sequence 19, Appl	242	6	1.5	35	4	US-09-834-784-164	Sequence 164, Appl
170	6	1.5	35	2	US-08-874-347-8	Sequence 8, Appl	243	6	1.5	35	4	US-09-834-784-165	Sequence 165, Appl
171	6	1.5	35	3	US-08-486-099-49	Sequence 49, Appl	244	6	1.5	37	4	US-09-255-208A-19	Sequence 19, Appl
172	6	1.5	35	3	US-08-486-099-147	Sequence 147, App	245	6	1.5	44	6	5188961-4	Patent No. 5188961
173	6	1.5	35	3	US-09-093-522-8	Sequence 8, Appl	246	6	1.5	46	2	US-09-057-762-15	Sequence 15, Appl

247	6	1.5	46	3	US-08-326-119A-15	Sequence 15, Appl	320	6	1.5	152	1	US-08-318-193-84	Sequence 84, Appl
248	6	1.5	48	1	US-07-796-361A-15	Sequence 15, Appl	321	6	1.5	152	4	US-09-198-452A-710	Sequence 110, App
249	6	1.5	59	1	US-08-650-528-7	Sequence 7, Appl	322	6	1.5	155	4	US-09-071-035-112	Sequence 112, App
250	6	1.5	59	3	US-09-060-584-7	Sequence 7, Appl	323	6	1.5	159	4	US-09-252-991A-28211	Sequence 28211, A
251	6	1.5	59	3	US-09-413-140A-7	Sequence 7, Appl	324	6	1.5	162	4	US-09-732-210-1445	Sequence 1445, Ap
252	6	1.5	60	6	5188961-1	Patent No. 5188961	325	6	1.5	163	4	US-09-252-991A-18617	Sequence 18617, Ap
253	6	1.5	61	4	US-09-205-258-260	Sequence 260, App	326	6	1.5	163	4	US-09-252-991A-20974	Sequence 20974, A
254	6	1.5	64	4	US-09-687-698-22	Sequence 403, App	327	6	1.5	163	4	US-09-252-991A-20971	Sequence 20971, A
255	6	1.5	65	3	US-08-905-223-403	Sequence 403, App	328	6	1.5	165	4	US-09-732-210-1608	Sequence 1608, Ap
256	6	1.5	66	4	US-09-134-001C-4566	Sequence 4566, Ap	329	6	1.5	166	4	US-09-134-001C-3657	Sequence 3657, Ap
257	6	1.5	70	3	US-08-486-099-100	Sequence 100, App	330	6	1.5	166	4	US-09-252-991A-18554	Sequence 18554, A
258	6	1.5	70	3	US-08-360-107A-91	Sequence 91, Appl	331	6	1.5	166	4	US-09-252-991A-22139	Sequence 22139, Ap
259	6	1.5	70	3	US-08-360-107A-110	Sequence 110, App	332	6	1.5	167	4	US-09-134-001C-2867	Sequence 2867, Ap
260	6	1.5	70	3	US-08-484-221B-100	Sequence 100, App	333	6	1.5	169	4	US-09-328-352-4491	Sequence 4491, Ap
261	6	1.5	70	3	US-08-919-597-100	Sequence 100, App	334	6	1.5	170	2	US-08-483-101-3	Sequence 3, Appl
262	6	1.5	70	3	US-08-475-668A-100	Sequence 100, App	335	6	1.5	172	2	US-08-853-659A-46	Sequence 46, Appl
263	6	1.5	70	3	US-08-485-551A-100	Sequence 100, App	336	6	1.5	173	4	US-08-671-548C-50	Sequence 50, Appl
264	6	1.5	70	3	US-08-471-913A-100	Sequence 100, App	337	6	1.5	175	4	US-08-687-688-74	Sequence 24, Appl
265	6	1.5	70	3	US-08-485-264A-100	Sequence 100, App	338	6	1.5	177	1	US-08-284-393B-7	Sequence 7, Appl
266	6	1.5	70	4	US-08-474-349A-100	Sequence 100, App	339	6	1.5	177	1	US-08-446-908-8	Sequence 4, Appl
267	6	1.5	70	4	US-08-255-208A-36	Sequence 36, Appl	340	6	1.5	177	1	US-08-231-205A-4	Sequence 4, Appl
268	6	1.5	70	4	US-08-470-896-100	Sequence 100, App	341	6	1.5	177	2	US-08-850-910A-18	Sequence 18, Appl
269	6	1.5	70	4	US-08-485-546A-100	Sequence 100, App	342	6	1.5	177	2	US-08-871-161-4	Sequence 4, Appl
270	6	1.5	72	4	US-08-918-428D-4	Sequence 4, Appl	343	6	1.5	177	4	US-08-469-280A-49	Sequence 49, Appl
271	6	1.5	75	3	US-08-301-162-6	Sequence 6, Appl	344	6	1.5	177	4	US-08-671-548C-42	Sequence 42, Appl
272	6	1.5	75	4	US-09-461-240-6	Sequence 6, Appl	345	6	1.5	177	4	US-08-671-548C-44	Sequence 44, Appl
273	6	1.5	75	4	US-09-968-927-6	Sequence 6, Appl	346	6	1.5	177	4	US-08-488-446-49	Sequence 49, Appl
274	6	1.5	77	3	US-09-100-802-3	Sequence 3, Appl	347	6	1.5	177	4	US-08-467-344A-49	Sequence 49, Appl
275	6	1.5	78	4	US-09-732-210-820	Sequence 820, App	348	6	1.5	177	5	PCT-US95-08950-7	Sequence 7, Appl
276	6	1.5	79	4	US-09-198-452A-676	Sequence 676, App	349	6	1.5	179	4	US-08-874-102-41	Sequence 41, Appl
277	6	1.5	85	4	US-09-134-001C-5400	Sequence 5400, Ap	350	6	1.5	179	4	US-08-874-102-44	Sequence 44, Appl
278	6	1.5	85	4	US-09-328-352-5627	Sequence 5627, Ap	351	6	1.5	179	4	US-08-984-919A-41	Sequence 41, Appl
279	6	1.5	85	4	US-09-732-210-1469	Sequence 1469, Ap	352	6	1.5	179	4	US-08-984-919A-41	Sequence 44, Appl
280	6	1.5	86	4	US-08-311-731A-242	Sequence 242, App	353	6	1.5	179	4	US-09-252-991A-22072	Sequence 22072, A
281	6	1.5	92	4	US-09-252-991A-16632	Sequence 16632, A	354	6	1.5	179	4	US-09-107-532A-3794	Sequence 3794, Ap
282	6	1.5	94	4	US-08-946-329A-88	Sequence 88, Appl	355	6	1.5	182	4	US-09-252-991A-27281	Sequence 27281, A
283	6	1.5	96	2	US-08-668-255-7	Sequence 7, Appl	356	6	1.5	184	4	US-09-252-991A-22743	Sequence 22743, A
284	6	1.5	96	2	US-08-668-255-9	Sequence 9, Appl	357	6	1.5	185	4	US-09-252-991A-20685	Sequence 20685, A
285	6	1.5	103	2	US-09-252-991A-27103	Sequence 27103, A	358	6	1.5	185	4	US-09-252-991A-23490	Sequence 23490, A
286	6	1.5	105	2	US-08-668-255-5	Sequence 5, Appl	359	6	1.5	186	1	US-08-026-758-25	Sequence 25, Appl
287	6	1.5	105	4	US-09-252-991A-25169	Sequence 25169, A	360	6	1.5	186	1	US-08-026-758-26	Sequence 26, Appl
288	6	1.5	108	4	US-09-252-991A-19573	Sequence 19573, A	361	6	1.5	186	4	US-09-487-792-15	Sequence 15, Appl
289	6	1.5	108	4	US-09-732-210-776	Sequence 776, App	362	6	1.5	186	4	US-09-908-594-15	Sequence 15, Appl
290	6	1.5	111	4	US-09-732-210-778	Sequence 778, App	363	6	1.5	187	4	US-09-252-991A-26269	Sequence 26269, A
291	6	1.5	115	4	US-09-328-352-4375	Sequence 4375, Ap	364	6	1.5	189	4	US-09-082-358B-8	Sequence 8, Appl
292	6	1.5	121	4	US-09-775-932-14	Sequence 14, Appl	365	6	1.5	189	4	US-08-671-548C-48	Sequence 48, Appl
293	6	1.5	122	4	US-09-732-210-241	Sequence 241, App	366	6	1.5	189	4	US-08-671-548C-52	Sequence 52, Appl
294	6	1.5	122	4	US-09-732-210-585	Sequence 585, App	367	6	1.5	190	1	US-08-106-981-2	Sequence 2, Appl
295	6	1.5	123	4	US-09-149-476-693	Sequence 693, App	368	6	1.5	190	4	US-08-918-428D-3	Sequence 3, Appl
296	6	1.5	126	2	US-08-675-508-20	Sequence 20, Appl	369	6	1.5	190	4	US-09-252-991A-18880	Sequence 18880, A
297	6	1.5	127	4	US-08-311-731A-226	Sequence 226, App	370	6	1.5	192	4	US-08-671-548C-32	Sequence 32, Appl
298	6	1.5	128	4	US-09-775-932-12	Sequence 12, Appl	371	6	1.5	193	4	US-08-671-548C-40	Sequence 40, Appl
299	6	1.5	129	4	US-09-732-210-282	Sequence 282, App	372	6	1.5	193	4	US-09-252-991A-17623	Sequence 17623, A
300	6	1.5	134	4	US-09-252-991A-25959	Sequence 25959, A	373	6	1.5	193	4	US-09-252-991A-29167	Sequence 29167, A
301	6	1.5	136	4	US-09-370-838-123	Sequence 123, App	374	6	1.5	194	4	US-09-071-035-110	Sequence 110, App
302	6	1.5	139	4	US-09-252-991A-24414	Sequence 24414, A	375	6	1.5	194	4	US-09-252-991A-31239	Sequence 31239, A
303	6	1.5	141	4	US-09-732-210-1330	Sequence 1330, Ap	376	6	1.5	196	1	US-08-063-552-7	Sequence 7, Appl
304	6	1.5	143	4	US-09-252-991A-23960	Sequence 23960, A	377	6	1.5	196	4	US-09-134-001C-4732	Sequence 4732, Ap
305	6	1.5	143	4	US-09-328-352-4829	Sequence 4829, Ap	378	6	1.5	196	5	PCT-US93-05704-7	Sequence 7, Appl
306	6	1.5	146	2	US-08-619-708A-4	Sequence 4, Appl	379	6	1.5	197	4	US-09-252-991A-31218	Sequence 31218, A
307	6	1.5	146	2	US-09-353-719-1	PCT-US95-07135-2	380	6	1.5	197	4	US-09-252-991A-23002	Sequence 23002, A
308	6	1.5	148	5	PCT-US95-07135-2	Sequence 2, Appl	381	6	1.5	199	4	US-09-252-991A-26549	Sequence 26549, A
309	6	1.5	149	2	US-08-461-030C-2	Sequence 2, Appl	382	6	1.5	200	4	US-09-107-532A-5925	Sequence 5925, Ap
310	6	1.5	149	3	US-08-744-138-2	Sequence 2, Appl	383	6	1.5	201	1	US-08-444-083-8	Sequence 8, Appl
311	6	1.5	149	3	US-09-431-480-8	Sequence 8, Appl	384	6	1.5	201	1	US-08-286-504-8	Sequence 8, Appl
312	6	1.5	149	3	US-09-431-480-10	Sequence 10, Appl	385	6	1.5	201	1	US-08-442-745-8	Sequence 8, Appl
313	6	1.5	149	3	US-09-617-302-8	Sequence 8, Appl	386	6	1.5	201	1	US-08-443-129-8	Sequence 8, Appl
314	6	1.5	149	3	US-09-617-302-10	Sequence 10, Appl	387	6	1.5	201	1	US-08-443-129-8	Sequence 8, Appl
315	6	1.5	149	4	US-09-241-376-2	Sequence 2, Appl	388	6	1.5	201	1	US-08-443-130-8	Sequence 8, Appl
316	6	1.5	150	4	US-09-252-991A-17264	Sequence 17264, A	389	6	1.5	201	1	US-08-792-019B-11	Sequence 11, Appl
317	6	1.5	151	4	US-09-252-991A-27007	Sequence 27007, A	390	6	1.5	201	3	US-09-106-182-4	Sequence 4, Appl
318	6	1.5	151	4	US-09-732-210-1297	Sequence 1297, Ap	391	6	1.5	201	3	US-08-988-819-11	Sequence 11, Appl
319	6	1.5	151	6	5229115-1	Patent No. 5229115	392	6	1.5	201	3	US-08-898-911-8	Sequence 8, Appl

393	6	1.5	201	3	US-09-016-534-11	Sequence 11, Appl	466	6	1.5	248	3	US-09-417-090-26	Sequence 26, Appl
394	6	1.5	201	4	US-09-648-183-3	Sequence 3, Appl	467	6	1.5	248	3	US-09-189-060B-2	Sequence 2, Appl
395	6	1.5	201	5	PCT-US95-04457-8	Sequence 8, Appl	468	6	1.5	248	3	US-09-349-677-6	Sequence 6, Appl
396	6	1.5	202	4	US-09-252-991A-23766	Sequence 23766, A	469	6	1.5	248	4	US-08-470-953A-6	Sequence 6, Appl
397	6	1.5	203	4	US-09-252-991A-19706	Sequence 19706, A	470	6	1.5	248	4	US-09-727-578-22	Sequence 22, Appl
398	6	1.5	204	4	US-09-323-872A-13	Sequence 13, Appl	471	6	1.5	248	4	US-09-727-578-24	Sequence 24, Appl
399	6	1.5	204	4	US-09-149-476-429	Sequence 429, App	472	6	1.5	248	4	US-09-727-578-26	Sequence 26, Appl
400	6	1.5	204	4	US-09-325-932A-145	Sequence 145, App	473	6	1.5	249	3	US-08-750-155A-3	Sequence 3, Appl
401	6	1.5	204	4	US-09-252-991A-23428	Sequence 23428, A	474	6	1.5	249	3	US-08-975-698A-3	Sequence 3, Appl
402	6	1.5	204	4	US-09-252-991A-30470	Sequence 30470, A	475	6	1.5	249	3	US-09-417-090-3	Sequence 3, Appl
403	6	1.5	204	4	US-09-072-433-22	Sequence 22, Appl	476	6	1.5	249	4	US-09-727-578-3	Sequence 3, Appl
404	6	1.5	205	4	US-09-252-991A-25068	Sequence 25068, A	477	6	1.5	250	4	US-09-252-991A-25449	Sequence 25449, A
405	6	1.5	205	4	US-09-252-991A-28196	Sequence 28196, A	478	6	1.5	253	4	US-09-149-476-576	Sequence 576, App
406	6	1.5	206	1	US-08-709-912-7	Sequence 7, Appl	479	6	1.5	253	4	US-09-198-452A-385	Sequence 385, App
407	6	1.5	206	2	US-09-047-370-7	Sequence 7, Appl	480	6	1.5	254	4	US-09-252-991A-19511	Sequence 19511, A
408	6	1.5	206	4	US-09-107-532A-3801	Sequence 3801, Ap	481	6	1.5	254	4	US-09-252-991A-25583	Sequence 25583, A
409	6	1.5	207	3	US-08-235-836C-36	Sequence 36, Appl	482	6	1.5	255	2	US-08-310-912A-106	Sequence 106, App
410	6	1.5	207	3	US-09-185-501B-14	Sequence 14, Appl	483	6	1.5	255	3	US-08-841-089-106	Sequence 106, App
411	6	1.5	209	4	US-09-795-926-37	Sequence 37, Appl	484	6	1.5	255	3	PCT-US95-04570-106	Sequence 106, App
412	6	1.5	211	4	US-09-252-991A-29679	Sequence 29679, A	485	6	1.5	255	5	PCT-US95-04589-106	Sequence 106, App
413	6	1.5	212	2	US-08-761-248B-2	Sequence 2, Appl	486	6	1.5	255	2	US-08-658-277-2	Sequence 2, Appl
414	6	1.5	212	2	US-08-761-248B-6	Sequence 6, Appl	487	6	1.5	256	2	US-08-658-277-3	Sequence 3, Appl
415	6	1.5	212	4	US-09-199-637A-235	Sequence 235, App	488	6	1.5	256	2	US-09-107-532A-3657	Sequence 3657, Ap
416	6	1.5	213	1	US-07-930-678-2	Sequence 3, Appl	489	6	1.5	257	4	US-09-328-352-6231	Sequence 6231, Ap
417	6	1.5	213	3	US-08-943-173-3	Sequence 3, Appl	490	6	1.5	258	3	US-09-068-140A-6	Sequence 6, Appl
418	6	1.5	213	4	US-09-252-991A-21832	Sequence 21832, A	491	6	1.5	259	4	US-09-252-991A-28679	Sequence 28679, A
419	6	1.5	214	4	US-09-257-583-15	Sequence 15, Appl	492	6	1.5	261	1	US-08-245-688-2	Sequence 2, Appl
420	6	1.5	217	4	US-09-252-991A-22573	Sequence 22573, A	493	6	1.5	261	1	US-08-245-688-4	Sequence 4, Appl
421	6	1.5	217	4	US-09-252-991A-27623	Sequence 27623, A	494	6	1.5	261	1	US-08-245-688-6	Sequence 6, Appl
422	6	1.5	218	4	US-09-252-991A-29586	Sequence 29586, A	495	6	1.5	261	1	US-08-245-688-8	Sequence 8, Appl
423	6	1.5	221	1	US-08-698-978-2	Sequence 2, Appl	496	6	1.5	261	1	US-08-245-688-10	Sequence 10, Appl
424	6	1.5	221	4	US-08-470-953A-3	Sequence 3, Appl	497	6	1.5	261	1	US-08-245-688-12	Sequence 12, Appl
425	6	1.5	221	4	US-09-252-991A-19158	Sequence 19158, A	498	6	1.5	261	2	US-07-857-224B-58	Sequence 58, Appl
426	6	1.5	222	4	US-08-035-928-21	Sequence 21, Appl	499	6	1.5	261	3	US-08-768-373-2	Sequence 2, Appl
427	6	1.5	224	4	US-09-107-532A-6161	Sequence 6161, Ap	500	6	1.5	261	3	US-09-252-991A-18958	Sequence 18958, A
428	6	1.5	226	3	US-09-195-286-1	Sequence 1, Appl	501	6	1.5	263	4	US-09-702-705-325	Sequence 325, App
429	6	1.5	226	3	US-09-252-991A-11306	Sequence 11306, A	502	6	1.5	263	4	US-09-736-457-325	Sequence 325, App
430	6	1.5	228	3	US-09-303-064-24	Sequence 24, Appl	503	6	1.5	264	4	US-09-252-991A-18073	Sequence 2, Appl
431	6	1.5	228	4	US-09-086-503-24	Sequence 4, Appl	504	6	1.5	266	1	US-08-202-857-2	Sequence 2, Appl
432	6	1.5	229	3	US-08-750-145A-4	Sequence 4, Appl	505	6	1.5	266	1	US-09-589-287B-19	Sequence 19, Appl
433	6	1.5	229	3	US-08-975-698A-4	Sequence 4, Appl	506	6	1.5	266	4	US-09-879-919-24	Sequence 24, Appl
434	6	1.5	229	3	US-09-417-090-4	Sequence 4, Appl	507	6	1.5	266	4	US-09-588-991A-13609	Sequence 13609, A
435	6	1.5	229	4	US-09-727-578-4	Sequence 4, Appl	508	6	1.5	267	4	US-09-252-991A-30609	Sequence 4, Appl
436	6	1.5	229	4	US-09-252-991A-20272	Sequence 20272, A	509	6	1.5	269	1	US-08-202-857-4	Sequence 4, Appl
437	6	1.5	230	4	US-09-328-352-7924	Sequence 16763, A	510	6	1.5	269	1	US-09-082-649B-82	Sequence 82, Appl
438	6	1.5	231	4	US-09-489-847-176	Sequence 17624, Ap	511	6	1.5	269	4	US-09-252-991A-27504	Sequence 27504, A
439	6	1.5	233	4	US-09-101-135-2	Sequence 2, Appl	512	6	1.5	271	2	US-07-977-630-5	Sequence 5, Appl
440	6	1.5	235	3	US-09-303-064-27	Sequence 27, Appl	513	6	1.5	271	2	US-08-850-880-6	Sequence 6, Appl
441	6	1.5	236	4	US-09-086-503-27	Sequence 27, Appl	514	6	1.5	271	2	US-08-944-916-6	Sequence 6, Appl
442	6	1.5	236	4	US-09-252-991A-22298	Sequence 22298, A	515	6	1.5	271	2	US-08-814-877-6	Sequence 6, Appl
443	6	1.5	240	4	US-09-252-991A-21397	Sequence 21397, A	516	6	1.5	271	2	US-08-760-615-8	Sequence 8, Appl
444	6	1.5	240	4	US-09-795-926-33	Sequence 33, Appl	517	6	1.5	271	3	US-09-131-028A-5	Sequence 5, Appl
445	6	1.5	240	4	US-08-896-933-32	Sequence 32, Appl	518	6	1.5	271	3	US-09-414-432A-6	Sequence 6, Appl
446	6	1.5	241	4	US-09-314-235-32	Sequence 32, Appl	519	6	1.5	271	3	US-09-107-532A-7012	Sequence 7012, Ap
447	6	1.5	243	4	US-09-252-991A-18574	Sequence 18574, A	520	6	1.5	271	4	US-09-252-991A-22405	Sequence 22405, A
448	6	1.5	243	4	US-09-252-991A-28864	Sequence 28864, A	521	6	1.5	271	4	US-09-252-991A-22698	Sequence 22698, A
449	6	1.5	243	4	US-09-090-567-2	Sequence 5628, Ap	522	6	1.5	271	4	US-09-252-991A-22698	Sequence 22698, A
450	6	1.5	244	4	US-09-328-352-5628	Sequence 17982, A	523	6	1.5	273	4	US-09-252-991A-22405	Sequence 22405, A
451	6	1.5	245	4	US-09-252-991A-17982	Sequence 28029, A	524	6	1.5	280	4	US-09-252-991A-22698	Sequence 22698, A
452	6	1.5	245	4	US-09-252-991A-28029	Sequence 8, Appl	525	6	1.5	280	4	US-09-252-991A-22698	Sequence 22698, A
453	6	1.5	246	3	US-09-067-800-8	Sequence 19, Appl	526	6	1.5	282	4	US-09-107-532A-7012	Sequence 7012, Ap
454	6	1.5	246	3	US-09-349-677-8	Sequence 19, Appl	527	6	1.5	282	4	US-09-386-642-54	Sequence 54, Appl
455	6	1.5	247	3	US-08-501-126-19	Sequence 18, Appl	528	6	1.5	284	4	US-09-522-714-24	Sequence 24, Appl
456	6	1.5	248	3	US-08-750-145A-18	Sequence 20, Appl	529	6	1.5	285	4	US-09-286-529-1	Sequence 2, Appl
457	6	1.5	248	3	US-08-750-145A-20	Sequence 22, Appl	530	6	1.5	285	4	US-09-589-287B-2	Sequence 2, Appl
458	6	1.5	248	3	US-08-750-145A-22	Sequence 24, Appl	531	6	1.5	285	4	US-09-496-118B-1	Sequence 1, Appl
459	6	1.5	248	3	US-08-975-698A-22	Sequence 26, Appl	532	6	1.5	285	4	US-09-565-423-2	Sequence 2, Appl
460	6	1.5	248	3	US-08-975-698A-24	Sequence 22, Appl	533	6	1.5	285	4	US-09-879-919-23	Sequence 23, Appl
461	6	1.5	248	3	US-08-975-698A-26	Sequence 24, Appl	534	6	1.5	285	4	US-09-252-991A-18133	Sequence 18133, A
462	6	1.5	248	3	US-08-975-698A-28	Sequence 26, Appl	535	6	1.5	285	4		
463	6	1.5	248	3	US-09-067-800-6	Sequence 22, Appl	536	6	1.5	285	4		
464	6	1.5	248	3	US-09-417-090-22	Sequence 24, Appl							
465	6	1.5	248	3	US-09-417-090-24	Sequence 24, Appl							

539	6	1.5	285	4	US-09-588-947A-2	Sequence 2, Appli	612	6	1.5	332	4	US-09-252-991A-28700	Sequence 28700, A
540	6	1.5	287	2	US-08-901-306-2	Sequence 2, Appli	613	6	1.5	333	2	US-08-896-083-2	Sequence 2, Appli
541	6	1.5	287	3	US-09-180-271-2	Sequence 2, Appli	614	6	1.5	333	2	US-08-896-083-4	Sequence 4, Appli
542	6	1.5	287	4	US-09-252-991A-29951	Sequence 29951, A	615	6	1.5	333	3	US-09-110-910A-2	Sequence 2, Appli
543	6	1.5	288	4	US-09-386-642-13	Sequence 13, Appli	616	6	1.5	333	3	US-09-110-910A-4	Sequence 4, Appli
544	6	1.5	289	4	US-09-386-642-14	Sequence 14, Appli	617	6	1.5	334	2	US-08-484-397A-8	Sequence 8, Appli
545	6	1.5	289	4	US-09-252-991A-17482	Sequence 17482, A	618	6	1.5	334	6	5290690-11	Parent No. 5290690
546	6	1.5	289	4	US-08-311-731A-396	Sequence 396, App	619	6	1.5	335	4	US-09-252-991A-25335	Sequence 25335, A
547	6	1.5	290	4	US-09-328-352-7390	Sequence 7390, Ap	620	6	1.5	336	2	US-08-784-651-4	Sequence 4, Appli
548	6	1.5	291	4	US-09-252-991A-22826	Sequence 22826, A	621	6	1.5	336	2	US-08-997-080-156	Sequence 156, App
549	6	1.5	294	4	US-09-252-991A-37760	Sequence 37760, A	622	6	1.5	336	2	US-08-997-362-156	Sequence 156, App
550	6	1.5	294	4	US-09-328-352-4575	Sequence 4575, Ap	623	6	1.5	336	3	US-09-095-855-156	Sequence 156, App
551	6	1.5	295	4	US-09-134-001C-3589	Sequence 3589, Ap	624	6	1.5	336	3	US-09-105-390-52	Sequence 52, App
552	6	1.5	297	4	US-08-765-907A-3	Sequence 3, Appli	625	6	1.5	337	1	US-09-317-223-3	Sequence 60, Appli
553	6	1.5	298	4	US-09-252-991A-27658	Sequence 27658, A	626	6	1.5	337	1	US-08-317-223-3	Sequence 3, Appli
554	6	1.5	299	3	US-09-147-826B-2	Sequence 2, Appli	627	6	1.5	337	1	US-08-445-135-4	Sequence 4, Appli
555	6	1.5	299	4	US-09-252-991A-30704	Sequence 30704, A	628	6	1.5	337	3	US-09-055-849A-3	Sequence 3, Appli
556	6	1.5	300	4	US-09-328-352-5297	Sequence 5297, Ap	629	6	1.5	337	3	US-09-213-632-3	Sequence 3, Appli
557	6	1.5	301	4	US-09-252-991A-17405	Sequence 17405, A	630	6	1.5	337	4	US-09-195-637A-315	Sequence 315, App
558	6	1.5	302	4	US-08-887-534A-90	Sequence 90, Appli	631	6	1.5	337	4	US-09-252-991A-21400	Sequence 21400, A
559	6	1.5	302	4	US-09-527-431-90	Sequence 352, Appli	632	6	1.5	337	4	US-09-328-352-5787	Sequence 5787, Ap
560	6	1.5	303	4	US-09-134-001C-3177	Sequence 3177, Ap	633	6	1.5	337	5	PCT-US95-12675-3	Sequence 3, Appli
561	6	1.5	303	4	US-09-674-741-8	Sequence 8, Appli	634	6	1.5	338	4	US-09-328-352-7151	Sequence 7151, Ap
562	6	1.5	303	4	US-09-107-532A-4487	Sequence 4487, Ap	635	6	1.5	340	3	US-09-180-783-2	Sequence 2, Appli
563	6	1.5	304	4	US-09-252-991A-23116	Sequence 23116, A	636	6	1.5	342	4	US-09-252-991A-27223	Sequence 27223, A
564	6	1.5	306	3	US-09-105-390-36	Sequence 36, Appli	637	6	1.5	342	4	US-08-282-197C-57	Sequence 57, Appli
565	6	1.5	306	4	US-09-386-642-53	Sequence 53, Appli	638	6	1.5	343	2	US-08-282-197C-57	Sequence 57, Appli
566	6	1.5	306	4	US-09-252-991A-21754	Sequence 21754, A	639	6	1.5	344	4	US-09-724-623-79	Sequence 79, Appli
567	6	1.5	308	4	US-09-252-991A-18422	Sequence 18422, A	640	6	1.5	344	4	US-09-252-991A-16756	Sequence 16756, A
568	6	1.5	309	1	US-08-236-918A-2	Sequence 2, Appli	641	6	1.5	344	4	US-09-252-991A-19273	Sequence 19273, A
569	6	1.5	309	4	US-09-150-864A-2	Sequence 2, Appli	642	6	1.5	346	1	US-09-366-953A-45	Sequence 16732, A
570	6	1.5	311	3	US-09-105-390-44	Sequence 44, Appli	643	6	1.5	348	2	US-08-484-397A-2	Sequence 2, Appli
571	6	1.5	311	4	US-09-252-991A-22406	Sequence 22406, A	644	6	1.5	348	2	US-08-484-397A-3	Sequence 3, Appli
572	6	1.5	312	4	US-09-252-991A-31029	Sequence 31029, A	645	6	1.5	348	2	US-08-484-397A-4	Sequence 4, Appli
573	6	1.5	313	2	US-08-284-465-5	Sequence 5, Appli	646	6	1.5	348	2	US-08-484-397A-5	Sequence 5, Appli
574	6	1.5	313	2	US-08-770-544-10	Sequence 10, Appli	647	6	1.5	348	2	US-08-484-397A-6	Sequence 6, Appli
575	6	1.5	313	4	US-09-252-991A-32017	Sequence 32017, A	648	6	1.5	348	2	US-08-484-397A-7	Sequence 7, Appli
576	6	1.5	313	4	US-09-551-826D-14	Sequence 14, Appli	649	6	1.5	348	2	US-08-484-397A-27	Sequence 27, Appli
577	6	1.5	313	4	US-09-579-259-10	Sequence 10, Appli	650	6	1.5	348	2	US-08-484-397A-38	Sequence 38, Appli
578	6	1.5	314	4	US-09-328-352-5492	Sequence 5492, Ap	651	6	1.5	349	2	US-09-300-971A-9	Sequence 9, Appli
579	6	1.5	315	4	US-09-386-653A-9	Sequence 9, Appli	652	6	1.5	351	4	US-09-252-991A-18476	Sequence 18476, A
580	6	1.5	315	4	US-09-252-991A-20012	Sequence 20012, A	653	6	1.5	351	4	US-09-252-991A-18590	Sequence 18590, A
581	6	1.5	316	4	US-09-387-375-9	Sequence 9, Appli	654	6	1.5	352	4	US-08-313-412-1	Sequence 1, Appli
582	6	1.5	317	3	US-08-949-155-6	Sequence 6, Appli	655	6	1.5	353	3	US-09-457-342-3	Sequence 3, Appli
583	6	1.5	317	4	US-09-819-964-6	Sequence 6, Appli	656	6	1.5	353	4	US-09-328-352-8199	Sequence 8199, Ap
584	6	1.5	317	4	US-09-252-991A-18820	Sequence 18820, A	657	6	1.5	355	4	US-09-252-991A-33034	Sequence 33034, A
585	6	1.5	318	3	US-08-680-506-3	Sequence 3, Appli	658	6	1.5	357	4	US-09-252-991A-32582	Sequence 32582, A
586	6	1.5	318	4	US-09-328-352-6669	Sequence 6669, Ap	659	6	1.5	359	4	US-09-266-965-120	Sequence 120, App
587	6	1.5	319	4	US-09-386-642-12	Sequence 12, Appli	660	6	1.5	359	4	US-09-252-991A-23051	Sequence 23051, A
588	6	1.5	319	4	US-09-252-991A-23561	Sequence 23561, A	661	6	1.5	359	4	US-09-252-991A-28730	Sequence 28730, A
589	6	1.5	320	4	US-09-526-400-2	Sequence 2, Appli	662	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
590	6	1.5	321	4	US-09-252-991A-26003	Sequence 26003, A	663	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
591	6	1.5	321	4	US-09-795-926-35	Sequence 35, Appli	664	6	1.5	361	1	US-08-258-261B-3	Sequence 3, Appli
592	6	1.5	322	4	US-09-252-991A-18724	Sequence 18724, A	665	6	1.5	361	1	US-08-456-837-3	Sequence 3, Appli
593	6	1.5	323	4	US-09-252-991A-17111	Sequence 17111, A	666	6	1.5	361	1	US-08-457-342-3	Sequence 3, Appli
594	6	1.5	324	4	US-09-252-991A-27080	Sequence 27080, A	667	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
595	6	1.5	325	4	US-09-252-991A-28632	Sequence 28632, A	668	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
596	6	1.5	327	2	US-08-611-107-12	Sequence 12, Appli	669	6	1.5	361	1	US-08-457-646A-3	Sequence 3, Appli
597	6	1.5	327	2	US-08-422-560A-12	Sequence 12, Appli	670	6	1.5	361	2	US-08-729-214-3	Sequence 3, Appli
598	6	1.5	327	3	US-09-173-581-5	Sequence 5, Appli	671	6	1.5	361	3	US-09-028-934-3	Sequence 3, Appli
599	6	1.5	327	3	US-08-468-793-12	Sequence 12, Appli	672	6	1.5	361	3	US-09-028-934-3	Sequence 3, Appli
600	6	1.5	327	3	US-09-420-915-5	Sequence 5, Appli	673	6	1.5	361	4	US-09-501-115-8	Sequence 8, Appli
601	6	1.5	327	3	US-09-386-629-8	Sequence 8, Appli	674	6	1.5	362	4	US-09-252-991A-26917	Sequence 26917, A
602	6	1.5	327	4	US-09-252-991A-17609	Sequence 17609, A	675	6	1.5	363	4	US-09-252-991A-20195	Sequence 20195, A
603	6	1.5	327	4	US-09-252-991A-18775	Sequence 18775, A	676	6	1.5	364	4	US-09-198-452A-423	Sequence 423, App
604	6	1.5	328	4	US-09-386-642-11	Sequence 11, Appli	677	6	1.5	366	4	US-08-984-919A-11	Sequence 11, Appli
605	6	1.5	328	4	US-09-252-991A-17729	Sequence 17729, A	678	6	1.5	366	4	US-08-545-573A-1	Sequence 1, Appli
606	6	1.5	330	4	US-09-134-001C-3582	Sequence 3582, Ap	679	6	1.5	367	4	US-09-795-926-39	Sequence 39, Appli
607	6	1.5	330	4	US-09-252-991A-19159	Sequence 19159, A	680	6	1.5	367	4	US-09-252-991A-19846	Sequence 19846, A
608	6	1.5	331	3	US-09-094-557-34	Sequence 34, Appli	681	6	1.5	367	4	US-09-252-991A-29289	Sequence 29289, A
609	6	1.5	331	4	US-09-134-001C-5254	Sequence 5254, Ap	682	6	1.5	368	3	US-08-781-420-11	Sequence 11, Appli
610	6	1.5	331	4	US-09-252-991A-19406	Sequence 19406, A	683	6	1.5	368	4	US-08-874-102-11	Sequence 11, Appli
611	6	1.5	331	4	US-09-252-991A-29393	Sequence 29393, A	684	6	1.5	368	4	US-09-006-595A-11	Sequence 11, Appli

685	6	1.5	369	1	US-08-700-359-21	Sequence 21, Appl	758	6	1.5	405	4	US-09-252-991A-23838	Sequence 23838, A
686	6	1.5	369	4	US-09-252-991A-17960	Sequence 17960, A	759	6	1.5	406	4	US-09-198-452A-387	Sequence 387, App
687	6	1.5	369	4	US-08-311-731A-38	Sequence 38, Appl	760	6	1.5	409	4	US-09-613-303-55	Sequence 35, Appl
688	6	1.5	370	4	US-09-252-991A-21152	Sequence 21152, A	761	6	1.5	409	4	US-09-328-352-457	Sequence 4570, Ap
689	6	1.5	370	4	US-08-311-731A-203	Sequence 203, App	762	6	1.5	410	4	US-09-252-991A-3050	Sequence 30590, A
690	6	1.5	372	4	US-09-107-532A-4779	Sequence 4779, Ap	763	6	1.5	412	4	US-09-252-991A-22952	Sequence 22952, A
691	6	1.5	373	3	US-08-746-883-4	Sequence 4, Appl	764	6	1.5	414	3	US-09-334-601-4	Sequence 4, Appl
692	6	1.5	373	3	US-09-188-579-113	Sequence 113, App	765	6	1.5	414	4	US-09-252-991A-24714	Sequence 24714, A
693	6	1.5	373	3	US-09-315-444-113	Sequence 113, App	766	6	1.5	416	4	US-09-252-991A-18947	Sequence 18947, A
694	6	1.5	373	3	US-09-721-362-113	Sequence 113, App	767	6	1.5	417	4	US-09-599-360B-88	Sequence 88, Appl
695	6	1.5	375	1	US-08-121-714-5	Sequence 5, Appl	768	6	1.5	418	4	US-09-252-991A-17826	Sequence 17826, A
696	6	1.5	375	1	US-08-203-719-2	Sequence 2, Appl	769	6	1.5	419	3	US-09-011-197-4	Sequence 4, Appl
697	6	1.5	375	1	US-08-477-108A-5	Sequence 5, Appl	770	6	1.5	420	3	US-09-252-991A-50920	Sequence 30920, A
698	6	1.5	375	2	US-08-477-112-5	Sequence 5, Appl	771	6	1.5	421	3	US-09-198-452A-932	Sequence 932, App
699	6	1.5	375	2	US-08-746-883-5	Sequence 5, Appl	772	6	1.5	423	2	US-08-715-554-2	Sequence 2, Appl
700	6	1.5	375	5	PCR-US93-08332-5	Sequence 5, Appl	773	6	1.5	423	2	US-08-583-118-2	Sequence 2, Appl
701	6	1.5	375	4	US-09-252-991A-20454	Sequence 20454, A	774	6	1.5	423	4	US-09-073-009-142	Sequence 142, App
702	6	1.5	377	2	US-08-455-968B-3	Sequence 3, Appl	775	6	1.5	424	4	US-09-252-991A-17882	Sequence 17882, A
703	6	1.5	377	4	US-09-410-464-8	Sequence 8, Appl	776	6	1.5	424	4	US-09-328-352-4199	Sequence 4199, Ap
704	6	1.5	377	4	US-09-252-991A-19827	Sequence 19827, A	777	6	1.5	426	4	US-09-327-487A-1	Sequence 1, Appl
705	6	1.5	378	2	US-08-455-968B-10	Sequence 10, Appl	778	6	1.5	426	4	US-09-327-487A-5	Sequence 5, Appl
706	6	1.5	378	2	US-08-823-516-138	Sequence 138, App	779	6	1.5	426	4	US-09-327-487A-6	Sequence 6, Appl
707	6	1.5	378	4	US-09-252-991A-23047	Sequence 23047, A	780	6	1.5	426	4	US-09-327-487A-7	Sequence 7, Appl
708	6	1.5	378	4	US-09-328-352-8178	Sequence 8178, Ap	781	6	1.5	426	4	US-09-071-035-2	Sequence 2, Appl
709	6	1.5	379	2	US-08-887-365-36	Sequence 36, Appl	782	6	1.5	427	4	US-09-134-001C-4803	Sequence 4803, Ap
710	6	1.5	379	4	US-09-252-991A-31643	Sequence 31643, A	783	6	1.5	428	4	US-09-134-001C-3801	Sequence 3801, Ap
711	6	1.5	380	2	US-08-455-968B-1	Sequence 1, Appl	784	6	1.5	428	4	US-09-252-991A-22337	Sequence 22337, A
712	6	1.5	380	2	US-08-823-516-137	Sequence 137, App	785	6	1.5	428	4	US-09-252-991A-25955	Sequence 25955, A
713	6	1.5	380	3	US-09-426-557-10	Sequence 10, Appl	786	6	1.5	430	3	US-08-486-099-95	Sequence 95, Appl
714	6	1.5	380	4	US-09-252-991A-32470	Sequence 32470, A	787	6	1.5	430	3	US-08-360-107A-105	Sequence 105, App
715	6	1.5	380	4	US-09-107-532A-3873	Sequence 3873, Ap	788	6	1.5	430	3	US-08-484-223B-95	Sequence 95, Appl
716	6	1.5	380	5	PCR-US91-02560-4	Sequence 4, Appl	789	6	1.5	430	3	US-08-919-597-95	Sequence 95, Appl
717	6	1.5	382	4	US-09-252-991A-23727	Sequence 23727, A	790	6	1.5	430	3	US-08-475-668A-95	Sequence 95, Appl
718	6	1.5	382	4	US-09-328-352-7748	Sequence 30714, A	791	6	1.5	430	3	US-08-485-551A-95	Sequence 95, Appl
719	6	1.5	382	4	US-09-328-352-7748	Sequence 7748, Ap	792	6	1.5	430	3	US-08-471-913A-95	Sequence 95, Appl
720	6	1.5	384	3	US-08-968-563-12	Sequence 12, Appl	793	6	1.5	430	3	US-08-485-264A-95	Sequence 95, Appl
721	6	1.5	384	3	US-08-969-683A-12	Sequence 12, Appl	794	6	1.5	430	4	US-08-474-349A-95	Sequence 95, Appl
722	6	1.5	384	4	US-09-297-928-8	Sequence 8, Appl	795	6	1.5	430	4	US-08-255-208A-31	Sequence 31, Appl
723	6	1.5	386	3	US-09-293-322C-5	Sequence 5, Appl	796	6	1.5	430	4	US-08-470-896-95	Sequence 95, Appl
724	6	1.5	386	4	US-09-325-932A-143	Sequence 143, App	797	6	1.5	430	4	US-08-485-546A-95	Sequence 95, Appl
725	6	1.5	386	4	US-08-545-573A-2	Sequence 2, Appl	798	6	1.5	430	4	US-09-252-991A-18400	Sequence 18400, A
726	6	1.5	386	4	US-08-545-573A-39	Sequence 39, Appl	799	6	1.5	431	4	US-09-107-532A-7079	Sequence 7079, Ap
727	6	1.5	386	4	US-09-839-497A-5	Sequence 5, Appl	800	6	1.5	432	4	US-09-328-352-4131	Sequence 4131, Ap
728	6	1.5	386	4	US-09-198-452A-388	Sequence 388, App	801	6	1.5	434	4	US-09-252-991A-30855	Sequence 30855, A
729	6	1.5	387	4	US-09-252-991A-18477	Sequence 18477, A	802	6	1.5	435	4	US-09-252-991A-17767	Sequence 17767, A
730	6	1.5	387	4	US-09-252-991A-22112	Sequence 22112, A	803	6	1.5	435	3	US-09-188-579-112	Sequence 112, App
731	6	1.5	390	4	US-09-634-238-331	Sequence 331, App	804	6	1.5	436	3	US-09-315-444-112	Sequence 112, App
732	6	1.5	392	4	US-09-647-224A-14	Sequence 14, Appl	805	6	1.5	436	4	US-09-721-362-112	Sequence 112, App
733	6	1.5	392	4	US-09-252-991A-29764	Sequence 29764, A	806	6	1.5	437	4	US-09-252-991A-24390	Sequence 24390, A
734	6	1.5	395	3	US-09-032-372-1	Sequence 1, Appl	807	6	1.5	438	4	US-09-198-452A-985	Sequence 985, App
735	6	1.5	396	2	US-08-878-989-16	Sequence 16, Appl	808	6	1.5	439	4	US-09-198-452A-986	Sequence 986, App
736	6	1.5	396	2	US-08-850-880-4	Sequence 4, Appl	809	6	1.5	442	4	US-09-252-991A-31743	Sequence 31743, A
737	6	1.5	396	2	US-08-944-916-4	Sequence 4, Appl	810	6	1.5	443	4	US-09-252-991A-18213	Sequence 18213, A
738	6	1.5	396	2	US-08-814-877-4	Sequence 4, Appl	811	6	1.5	443	4	US-09-252-991A-30036	Sequence 30036, A
739	6	1.5	396	3	US-09-108-020-38	Sequence 38, Appl	812	6	1.5	444	4	US-09-252-991A-32532	Sequence 32532, A
740	6	1.5	396	3	US-09-272-796-16	Sequence 16, Appl	813	6	1.5	444	4	US-09-252-991A-17767	Sequence 17767, A
741	6	1.5	396	3	US-09-344-700-2	Sequence 2, Appl	814	6	1.5	444	4	US-09-252-991A-24777	Sequence 24777, A
742	6	1.5	396	4	US-09-272-432A-4	Sequence 4, Appl	815	6	1.5	446	4	US-09-252-991A-21819	Sequence 21819, A
743	6	1.5	397	1	US-08-415-823-2	Sequence 2, Appl	816	6	1.5	448	4	US-09-252-991A-30230	Sequence 30230, A
744	6	1.5	397	2	US-09-086-662-2	Sequence 2, Appl	817	6	1.5	449	3	US-08-680-506-7	Sequence 7, Appl
745	6	1.5	397	2	US-08-311-731A-250	Sequence 250, App	818	6	1.5	449	3	US-09-252-991A-17164	Sequence 17164, A
746	6	1.5	398	2	US-08-599-171A-29	Sequence 29, Appl	819	6	1.5	450	4	US-09-252-991A-26266	Sequence 26266, A
747	6	1.5	398	2	US-08-646-590B-29	Sequence 29, Appl	820	6	1.5	450	4	US-09-369-247-97	Sequence 97, Appl
748	6	1.5	398	3	US-09-069-226-29	Sequence 29, Appl	821	6	1.5	451	4	US-09-252-991A-31986	Sequence 31986, A
749	6	1.5	398	3	US-09-412-184-29	Sequence 29, Appl	822	6	1.5	454	4	US-09-252-991A-18633	Sequence 18633, A
750	6	1.5	399	4	US-09-252-991A-16821	Sequence 16821, A	823	6	1.5	455	4	US-09-252-991A-31636	Sequence 31636, A
751	6	1.5	399	4	US-09-252-991A-20042	Sequence 20042, A	824	6	1.5	456	2	US-08-643-034A-4	Sequence 4, Appl
752	6	1.5	399	4	US-09-252-991A-23741	Sequence 23741, A	825	6	1.5	456	3	US-08-648-650A-4	Sequence 4, Appl
753	6	1.5	400	4	US-09-252-991A-24985	Sequence 24985, A	826	6	1.5	457	4	US-09-465-558-50	Sequence 50, Appl
754	6	1.5	401	4	US-09-252-991A-21672	Sequence 21672, A	827	6	1.5	457	3	US-09-183-266A-4	Sequence 4, Appl
755	6	1.5	403	2	US-08-846-762-89	Sequence 89, Appl	828	6	1.5	457	4	US-09-252-991A-19877	Sequence 19877, A
756	6	1.5	404	4	US-09-252-991A-24408	Sequence 24408, A	829	6	1.5	457	4	US-09-252-991A-22570	Sequence 22570, A
757	6	1.5	405	4	US-09-252-991A-23282	Sequence 23282, A	830	6	1.5	457	4		

831	6	1.5	457	4	US-08-311-731A-145	Sequence 145, App	904	6	1.5	501	3	US-09-307-185-6	Sequence 6, Appli
832	6	1.5	458	4	US-09-252-991A-20811	Sequence 20811, A	905	6	1.5	501	4	US-09-157-257-8	Sequence 8, Appli
833	6	1.5	458	4	US-09-252-991A-26855	Sequence 26855, A	906	6	1.5	501	4	US-09-252-991A-17730	Sequence 17730, A
834	6	1.5	460	4	US-09-252-991A-27768	Sequence 27768, A	907	6	1.5	502	4	US-09-252-991A-24388	Sequence 24388, A
835	6	1.5	460	4	US-09-252-991A-32153	Sequence 32153, A	908	6	1.5	503	4	US-09-252-991A-18052	Sequence 18052, A
836	6	1.5	461	1	US-07-796-361A-13	Sequence 13, Appli	909	6	1.5	503	4	US-09-252-991A-29343	Sequence 29343, A
837	6	1.5	462	1	US-07-612-673-4	Sequence 4, Appli	910	6	1.5	506	3	US-09-303-064-52	Sequence 52, Appli
838	6	1.5	462	1	US-08-539-666-4	Sequence 4, Appli	911	6	1.5	506	3	US-09-086-503-12	Sequence 52, Appli
839	6	1.5	462	4	US-09-252-991A-31372	Sequence 31372, A	912	6	1.5	506	4	US-09-252-991A-19050	Sequence 19050, A
840	6	1.5	462	4	US-09-328-352-7128	Sequence 7128, Ap	913	6	1.5	506	4	US-09-252-991A-19151	Sequence 19151, A
841	6	1.5	463	4	US-09-252-991A-30002	Sequence 30002, A	914	6	1.5	506	4	US-09-252-991A-32150	Sequence 32150, A
842	6	1.5	464	4	US-09-252-991A-26602	Sequence 26602, A	915	6	1.5	507	4	US-09-130-337A-25	Sequence 25, Appli
843	6	1.5	466	4	US-08-984-919A-33	Sequence 33, Appli	916	6	1.5	507	4	US-09-252-991A-23598	Sequence 23598, A
844	6	1.5	467	4	US-09-252-991A-29057	Sequence 29057, A	917	6	1.5	508	4	US-09-369-247-167	Sequence 167, App
845	6	1.5	468	4	US-08-874-102-33	Sequence 33, Appli	918	6	1.5	509	4	US-09-252-991A-32576	Sequence 32576, A
846	6	1.5	469	3	US-08-378-313-33	Sequence 33, Appli	919	6	1.5	510	3	US-08-246-489-2	Sequence 2, Appli
847	6	1.5	470	4	US-08-984-919A-55	Sequence 55, Appli	920	6	1.5	510	4	US-09-252-991A-25076	Sequence 25076, A
848	6	1.5	470	4	US-09-252-991A-20500	Sequence 20500, A	921	6	1.5	512	4	US-09-390-234-24	Sequence 24, Appli
849	6	1.5	471	4	US-09-252-991A-27633	Sequence 27633, A	922	6	1.5	512	4	US-09-716-865-14	Sequence 14, Appli
850	6	1.5	472	4	US-08-874-102-55	Sequence 55, Appli	923	6	1.5	514	4	US-09-252-991A-22124	Sequence 22124, A
851	6	1.5	472	4	US-09-252-991A-20558	Sequence 20558, A	924	6	1.5	514	2	US-08-762-106-8	Sequence 8, Appli
852	6	1.5	472	4	US-09-252-991A-22840	Sequence 22840, A	925	6	1.5	516	2	US-08-676-166A-2	Sequence 2, Appli
853	6	1.5	474	3	US-09-315-444-116	Sequence 116, App	926	6	1.5	516	3	US-08-745-404-2	Sequence 2, Appli
854	6	1.5	474	4	US-09-721-362-116	Sequence 116, App	927	6	1.5	516	3	US-09-320-774-8	Sequence 8, Appli
855	6	1.5	474	4	US-09-252-991A-21280	Sequence 21280, A	928	6	1.5	519	3	US-09-252-991A-30815	Sequence 30815, A
856	6	1.5	476	3	US-09-188-579-114	Sequence 114, App	929	6	1.5	520	4	US-09-252-991A-33049	Sequence 33049, A
857	6	1.5	476	3	US-09-315-444-114	Sequence 114, App	930	6	1.5	521	4	US-09-134-001C-42930	Sequence 42930, Ap
858	6	1.5	476	4	US-09-134-218-6	Sequence 6, Appli	931	6	1.5	521	4	US-09-252-991A-30925	Sequence 30925, A
859	6	1.5	476	4	US-09-721-362-114	Sequence 114, App	932	6	1.5	522	4	US-08-469-260A-606	Sequence 606, App
860	6	1.5	476	4	US-09-529-157-4	Sequence 4, Appli	933	6	1.5	522	4	US-08-468-446-606	Sequence 606, App
861	6	1.5	480	1	US-07-882-292-2	Sequence 2, Appli	934	6	1.5	522	4	US-08-467-344A-606	Sequence 606, App
862	6	1.5	480	2	US-08-468-812-5	Sequence 5, Appli	935	6	1.5	523	4	US-09-252-991A-23875	Sequence 23875, A
863	6	1.5	480	2	US-08-331-644-2	Sequence 2, Appli	936	6	1.5	523	4	US-09-252-991A-33001	Sequence 33001, A
864	6	1.5	480	4	US-08-590-563-5	Sequence 5, Appli	937	6	1.5	526	4	US-09-252-991A-28402	Sequence 28402, A
865	6	1.5	480	4	US-09-770-621-5	Sequence 5, Appli	938	6	1.5	526	4	US-09-252-991A-32984	Sequence 32984, A
866	6	1.5	480	5	PCT-US93-04102-2	Sequence 13, Appli	939	6	1.5	527	2	US-08-762-106-9	Sequence 9, Appli
867	6	1.5	483	3	US-08-431-517F-13	Sequence 13, Appli	940	6	1.5	527	3	US-09-320-774-9	Sequence 9, Appli
868	6	1.5	485	2	US-09-252-991A-29267	Sequence 29267, A	941	6	1.5	529	4	US-09-252-991A-18188	Sequence 18188, A
869	6	1.5	485	2	US-08-749-391-2	Sequence 2, Appli	942	6	1.5	529	4	US-09-252-991A-18188	Sequence 18188, A
870	6	1.5	485	3	US-09-390-200-2	Sequence 2, Appli	943	6	1.5	530	4	US-09-252-991A-21863	Sequence 21863, A
871	6	1.5	486	3	US-08-906-743-4	Sequence 4, Appli	944	6	1.5	531	4	US-09-252-991A-21965	Sequence 21965, A
872	6	1.5	487	3	US-09-594-193-13	Sequence 13, Appli	945	6	1.5	531	4	US-09-252-991A-22509	Sequence 22509, A
873	6	1.5	488	4	US-09-252-991A-19034	Sequence 19034, A	946	6	1.5	531	4	US-09-252-991A-31109	Sequence 31109, A
874	6	1.5	488	4	US-09-252-991A-24316	Sequence 24316, A	947	6	1.5	532	1	US-08-339-152A-32	Sequence 32, Appli
875	6	1.5	488	4	US-09-252-991A-29894	Sequence 29894, A	948	6	1.5	532	1	US-08-657-192-9	Sequence 9, Appli
876	6	1.5	489	1	US-09-252-991A-25740	Sequence 25740, A	949	6	1.5	532	3	US-08-523-373-7	Sequence 7, Appli
877	6	1.5	490	1	US-08-201-118-5	Sequence 5, Appli	950	6	1.5	532	3	US-08-948-564-10	Sequence 10, Appli
878	6	1.5	490	1	US-08-201-118-11	Sequence 11, Appli	951	6	1.5	533	4	US-09-252-991A-22953	Sequence 22953, A
879	6	1.5	490	2	US-08-238-821B-5	Sequence 5, Appli	952	6	1.5	533	4	US-09-252-991A-17757	Sequence 17757, A
880	6	1.5	490	2	US-08-238-821B-11	Sequence 11, Appli	953	6	1.5	535	4	US-09-252-991A-24728	Sequence 24728, A
881	6	1.5	490	5	PCT-US95-05744-5	Sequence 5, Appli	954	6	1.5	535	4	US-09-252-991A-25662	Sequence 25662, A
882	6	1.5	490	5	PCT-US95-05744-11	Sequence 11, Appli	955	6	1.5	537	1	US-08-657-192-15	Sequence 15, Appli
883	6	1.5	492	2	US-08-468-812-4	Sequence 4, Appli	956	6	1.5	539	2	US-08-467-963C-2	Sequence 2, Appli
884	6	1.5	492	2	US-08-468-812-7	Sequence 7, Appli	957	6	1.5	539	2	US-08-838-189D-2	Sequence 2, Appli
885	6	1.5	492	4	US-08-590-563-4	Sequence 4, Appli	958	6	1.5	539	3	US-08-852-344D-2	Sequence 2, Appli
886	6	1.5	492	4	US-08-590-563-7	Sequence 7, Appli	959	6	1.5	539	3	US-08-344-639E-2	Sequence 2, Appli
887	6	1.5	492	4	US-09-770-621-4	Sequence 4, Appli	960	6	1.5	539	3	US-08-467-969A-2	Sequence 2, Appli
888	6	1.5	492	4	US-09-770-621-7	Sequence 7, Appli	961	6	1.5	539	3	US-08-467-961A-2	Sequence 2, Appli
889	6	1.5	493	3	US-09-177-349-5	Sequence 5, Appli	962	6	1.5	539	3	US-08-001-554A-2	Sequence 6, Appli
890	6	1.5	493	4	US-09-252-991A-30722	Sequence 30722, A	963	6	1.5	539	4	US-09-328-352-5433	Sequence 5433, Ap
891	6	1.5	493	4	US-09-252-991A-32463	Sequence 32463, A	964	6	1.5	539	4	US-09-328-352-5433	Sequence 22219, A
892	6	1.5	494	4	US-09-252-991A-24099	Sequence 24099, A	965	6	1.5	540	4	US-09-328-352-4514	Sequence 4514, Ap
893	6	1.5	494	4	US-09-252-991A-32739	Sequence 32739, A	966	6	1.5	541	2	US-08-540-804-16	Sequence 16, Appli
894	6	1.5	495	4	US-08-984-919A-47	Sequence 47, Appli	967	6	1.5	541	2	US-08-540-804-16	Sequence 16, Appli
895	6	1.5	495	4	US-08-984-919A-16768	Sequence 16768, A	968	6	1.5	541	2	US-08-540-804-16	Sequence 16, Appli
896	6	1.5	495	4	US-08-874-102-47	Sequence 47, Appli	969	6	1.5	541	2	US-08-540-804-16	Sequence 16, Appli
897	6	1.5	498	4	US-08-511-485-13	Sequence 13, Appli	970	6	1.5	541	2	US-08-511-485-13	Sequence 13, Appli
898	6	1.5	498	4	US-09-134-001C-5233	Sequence 5233, Ap	971	6	1.5	542	4	US-09-252-991A-31091	Sequence 31091, A
899	6	1.5	498	4	US-09-201-936-13	Sequence 13, Appli	972	6	1.5	543	4	US-09-252-991A-20430	Sequence 20430, A
900	6	1.5	498	4	US-09-252-991A-17608	Sequence 17608, A	973	6	1.5	545	4	US-09-252-991A-17065	Sequence 17065, A
901	6	1.5	498	4	US-09-252-991A-20702	Sequence 20702, A	974	6	1.5	545	4	US-09-252-991A-33044	Sequence 33044, A
902	6	1.5	499	4	US-09-252-991A-23929	Sequence 23929, A	975	6	1.5	551	1	US-08-120-960-2	Sequence 2, Appli
903	6	1.5	501	2	US-08-980-060-6	Sequence 6, Appli	976	6	1.5	551	2	US-08-700-548-4	Sequence 4, Appli

977 6 1.5 551 4 US-09-347-878-9 Sequence 9, Appl
978 6 1.5 552 3 US-08-745-404-3 Sequence 3, Appl
979 6 1.5 553 2 US-08-663-566A-13 Sequence 13, Appl
980 6 1.5 553 2 US-08-484-575A-14 Sequence 14, Appl
981 6 1.5 553 2 US-08-023-610-13 Sequence 13, Appl
982 6 1.5 553 2 US-08-288-065A-13 Sequence 13, Appl
983 6 1.5 553 2 US-08-362-240A-13 Sequence 14, Appl
984 6 1.5 553 3 US-08-477-459-14 Sequence 14, Appl
985 6 1.5 553 3 US-08-479-869-14 Sequence 14, Appl
986 6 1.5 553 3 US-08-486-414-14 Sequence 14, Appl
987 6 1.5 553 3 US-08-804-372A-11 Sequence 11, Appl
988 6 1.5 553 4 US-09-252-991A-17984 Sequence 17984, A
989 6 1.5 553 4 US-09-252-991A-19864 Sequence 19864, A
990 6 1.5 553 5 PCT-US94-01826A-14 Sequence 14, Appl
991 6 1.5 553 5 PCT-US94-02252A-14 Sequence 13, Appl
992 6 1.5 553 5 PCT-US95-10245-13 Sequence 13, Appl
993 6 1.5 553 6 5310678-1 Patent No. 5310678
994 6 1.5 554 4 US-09-252-991A-23106 Sequence 23106, A
995 6 1.5 555 4 US-09-252-991A-28313 Sequence 28313, A
996 6 1.5 555 4 US-09-252-991A-28848 Sequence 28848, A
997 6 1.5 556 4 US-09-252-991A-24832 Sequence 24832, A
998 6 1.5 557 2 US-08-808-982-6 Sequence 6, Appl
999 6 1.5 557 3 US-09-306-902A-6 Sequence 6, Appl
1000 6 1.5 558 4 US-09-328-352-5580 Sequence 5580, Ap

ALIGNMENTS

RESULT 1
US-09-461-325-398
; Sequence 398, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 398
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-325-398

Query Match 2.4%; Score 10; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALAAVA 19
| | | | |
Db 104 RAALAAVA 113

RESULT 2
US-09-328-352-7383
; Sequence 7383, Application US/09328352

Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7383
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7383

Query Match 2.2%; Score 9; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 GDKVVEGI 369
| | | | |
Db 401 GDKVVEGI 409

RESULT 3
US-09-732-210-240
; Sequence 240, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitcanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 240
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-732-210-240

Query Match 1.9%; Score 8; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
| | | | |
Db 32 DKVVEGI 39

RESULT 4
US-09-732-210-764
; Sequence 764, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitcanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513

;; PRIOR FILING DATE: 1999-12-07
;; PRIOR APPLICATION NUMBER: US 60/169,340
;; PRIOR FILING DATE: 1999-12-07
;; NUMBER OF SEQ ID NOS: 1753
;; SEQ ID NO 764
;; LENGTH: 111
;; TYPE: PRT
;; ORGANISM: Chlamydia trachomatis
US-09-732-210-764

Query Match 1.9%; Score 8; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
|||
Db 32 DKVVEGI 39

RESULT 5
US-09-461-325-397
; Sequence 397, Application US/09461325A
; Patent No. 6475753

;; GENERAL INFORMATION:
;; APPLICANT: Ruben et al.
;; TITLE OF INVENTION: 94 Human Secreted Proteins
;; FILE REFERENCE: P2029P1
;; CURRENT APPLICATION NUMBER: US/09/461,325A
;; EARLIER APPLICATION NUMBER: PCT/US99/13418
;; EARLIER FILING DATE: 1999-06-15
;; EARLIER APPLICATION NUMBER: 60/089,507
;; EARLIER FILING DATE: 1998-06-16
;; EARLIER APPLICATION NUMBER: 60/089,508
;; EARLIER FILING DATE: 1998-06-16
;; EARLIER APPLICATION NUMBER: 60/089,509
;; EARLIER FILING DATE: 1998-06-16
;; EARLIER APPLICATION NUMBER: 60/089,510
;; EARLIER FILING DATE: 1998-06-16
;; EARLIER APPLICATION NUMBER: 60/090,112
;; EARLIER FILING DATE: 1998-06-22
;; EARLIER APPLICATION NUMBER: 60/090,113
;; EARLIER FILING DATE: 1998-06-22
;; NUMBER OF SEQ ID NOS: 532
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 397
;; LENGTH: 194
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-461-325-397

Query Match 1.9%; Score 8; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AALAAVA 19
|||
Db 176 AALAAVA 183

RESULT 6
US-09-252-991A-24458
; Sequence 24458, Application US/09252991A
; Patent No. 6551795

;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18

;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 24458
;; LENGTH: 343
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24458

Query Match 1.9%; Score 8; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALAAV 18
|||
Db 225 AALAAV 232

RESULT 7
US-09-040-799-3
; Sequence 3, Application US/09040799
; Patent No. 585820

;; GENERAL INFORMATION:
;; APPLICANT: CHANG, YIE-HWA
;; TITLE OF INVENTION: CLONE OF A NUCLEOTIDE SEQUENCE ENCODING
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
;; STREET: 7733 FORSYTH BLVD., SUITE 1400
;; CITY: ST. LOUIS
;; STATE: MO
;; COUNTRY: USA
;; ZIP: 63105
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/040,799
;; FILING DATE: 18-MAR-1998
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: HOLLAND, DONALD R.
;; REGISTRATION NUMBER: 35,197
;; REFERENCE/DOCKET NUMBER: 16153-4639
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 314-727-5188
;; TELEFAX: 314-727-6092
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 478 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-040-799-3

Query Match 1.9%; Score 8; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||
Db 235 NAGDTTVL 242

RESULT 8
US-09-093-448-1
; Sequence 1, Application US/09093448A
; Patent No. 6207704
;; GENERAL INFORMATION:

```

; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/093,448A
; CURRENT FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-093-448-1

Query Match          1.9%; Score 8; DB 3; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
        |||||
Db      235 NAGDTTVL 242

RESULT 9
US-09-093-448-2
; Sequence 2, Application US/09093448A
; Patent No. 6207704
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/093,448A
; CURRENT FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-093-448-2

Query Match          1.9%; Score 8; DB 3; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
        |||||
Db      235 NAGDTTVL 242

RESULT 10
US-09-093-448-3
; Sequence 3, Application US/09093448A
; Patent No. 6207704
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/093,448A
; CURRENT FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 478
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
US-09-093-448-3

Query Match          1.9%; Score 8; DB 3; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
        |||||
Db      235 NAGDTTVL 242

RESULT 11
US-09-813-555-1
; Sequence 1, Application US/09813555
; Patent No. 6566541
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-813-555-1

Query Match          1.9%; Score 8; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
        |||||
Db      235 NAGDTTVL 242

RESULT 12
US-09-813-555-2
; Sequence 2, Application US/09813555
; Patent No. 6566541
; GENERAL INFORMATION:
; APPLICANT: Liu, Jun O.
; APPLICANT: Griffith, Eric C.
; APPLICANT: Su, Zhuang
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
; FILE REFERENCE: 0492611-0346
; CURRENT APPLICATION NUMBER: US/09/813,555
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-813-555-2

Query Match          1.9%; Score 8; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      203 NAGDTTVL 210
        |||||
Db      235 NAGDTTVL 242

RESULT 13
```

US-09-813-555-3
Sequence 3, Application US/09813555
Patent No. 6566541
GENERAL INFORMATION:
APPLICANT: Liu, Jun O.
APPLICANT: Griffith, Eric C.
TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors
TITLE OF INVENTION: and Uses Thereof
FILE REFERENCE: 0492611-0346
CURRENT APPLICATION NUMBER: US/09/813,555
CURRENT FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 478
TYPE: PRT
ORGANISM: Homo sapiens
US-09-813-555-3

Query Match 1.9%; Score 8; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
Db 235 NAGDTTVL 242

RESULT 14

US-09-010-999-11
Sequence 11, Application US/09010999
Patent No. 6132976

GENERAL INFORMATION:

APPLICANT: Poole, Anthony R.
APPLICANT: Hollander, Anthony P.
APPLICANT: Billingham, R. C.
TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,999
FILING DATE: 22-JAN-1998
CLASSIFICATION: 4335
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/448,501
FILING DATE: 17-JUL-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/984,123
FILING DATE: 04-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 032931/0212
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-010-999-11

Query Match 1.7%; Score 7; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 IAGITGA 377
Db 11 IAGITGA 17

RESULT 15

US-09-227-357-347
Sequence 347, Application US/09227357
Patent No. 6342581

GENERAL INFORMATION:

APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P2010P1
CURRENT APPLICATION NUMBER: US/09/227,357
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684
EARLIER FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 347
LENGTH: 24
TYPE: PRT
ORGANISM: Homo sapiens
US-09-227-357-347

Query Match 1.7%; Score 7; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 NIESARA 117
DB 8 NIESARA 14

RESULT 16
US-09-198-452A-675
Sequence 675, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Grifffais, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 675
LENGTH: 111
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-675

Query Match 1.7%; Score 7; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEG 368
DB 32 DKVVEG 38

RESULT 17
US-08-312-387B-2
Sequence 2, Application US/08312387B
Patent No. 5545553
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:

ADDRESSER: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312,387B
FILING DATE: July 7, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-312-387B-2

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAAA 17
DB 13 AAAA 19

RESULT 18
US-08-683-426-2
Sequence 2, Application US/08683426
Patent No. 5705367
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,426
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B

TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-458-2

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAA 17
13 AAALAA 19

RESULT 19
US-08-683-458-2
Sequence 2, Application US/08683458
Patent No. 5798233

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,458
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-683-458-2

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAA 17
13 AAALAA 19

RESULT 20
US-08-878-360-2
Sequence 2, Application US/08878360
Patent No. 5945322

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:

APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-2

Query Match 1.7%; Score 7; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 AAALAA 17
13 AAALAA 19

RESULT 21
US-08-478-140B-2
Sequence 2, Application US/08478140B
Patent No. 6127153

GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-2

Query Match 1.7%; Score 7; DB 3; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAA 17
Db 13 AAALAA 19

RESULT 22
US-09-333-412-2
Sequence 2, Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-JUN-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-333-412-2

Query Match 1.7%; Score 7; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAA 17
Db 13 AAALAA 19

RESULT 23
US-09-338-943-2
Sequence 2, Application US/09338943
Patent No. 6379933
GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,943
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/478,140
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-338-943-2

Query Match 1.7%; Score 7; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAA 17

Db 13 AAALAAA 19

RESULT 24

US-08-664-596B-18
Sequence 18, Application US/08664596B

Patent No. 5807703

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John

APPLICANT: Lavallie, Edward

APPLICANT: Racie, Lisa

APPLICANT: Meiberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Evans, Cheryl

APPLICANT: Spaulding, Vikki

APPLICANT: Bowman, Michael

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

TITLE OF INVENTION: ENCODING THEM

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/664,596B

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 498-5851

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-664-596B-18

Query Match 1.7%; Score 7; DB 1; Length 131;

Best Local Similarity 100.0%; Pred. No. 46;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 SEAKTAS 409

Db 97 SEAKTAS 103

RESULT 25

US-08-534-975-4
Sequence 4, Application US/08534975

Patent No. 5723313

GENERAL INFORMATION:

APPLICANT: Sherr, Charles, J.

APPLICANT: Quelle, Dawn, B.

TITLE OF INVENTION: ARF-p19, A No. 5723313el Regulator of the Mammalian Cell

TITLE OF INVENTION: Cycle

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/534,975

FILING DATE: 28-SEP-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel, L.

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 0656.0590000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2540

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 132 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-534-975-4

Query Match 1.7%; Score 7; DB 1; Length 132;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVALVL 22

Db 41 AAVALVL 47

RESULT 26

US-08-954-470-4
Sequence 4, Application US/08954470

Patent No. 5876965

GENERAL INFORMATION:

APPLICANT: Sherr, Charles, J.

APPLICANT: Quelle, Dawn, B.

TITLE OF INVENTION: ARF-p19, A No. 5876965el Regulator of the

TITLE OF INVENTION: Mammalian Cell Cycle

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/954,470

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/534,975

FILING DATE: 28-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fox, Samuel, L.

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 0656.0590000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-954-470-4

Query Match 1.7%; Score 7; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 AAVAVL 22
|||||
Db 41 AAVAVL 47

RESULT 27
US-09-129-855A-4
Sequence 4, Application US/09129855A
Patent No. 6046032

GENERAL INFORMATION:
APPLICANT: Sherr, Charles, J.
APPLICANT: Quelle, Dawn, E.
TITLE OF INVENTION: ARF-p19, A No. 6046032el Regulator of the Mammalian
TITLE OF INVENTION: Cell Cycle
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/129,855A
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/954,470
FILING DATE: 20-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/534,975

FILING DATE: 27-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1340-1-023 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-129-855A-4

Query Match 1.7%; Score 7; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 AAVAVL 22
|||||
Db 41 AAVAVL 47

RESULT 28
US-09-247-154-4
Sequence 4, Application US/09247154
Patent No. 6172194

GENERAL INFORMATION:
APPLICANT: Sherr, Charles, J.
APPLICANT: Quelle, Dawn, E.
TITLE OF INVENTION: ARF-p19, A No. 6172194el Regulator of the
TITLE OF INVENTION: Mammalian Cell Cycle
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/247,154
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/954,470

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel, L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656.0590000

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-247-154-4

Query Match 1.7%; Score 7; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 AAVAVL 22
|||||
Db 41 AAVAVL 47

RESULT 29
US-09-480-718-4
Sequence 4, Application US/09480718
Patent No. 6407062

GENERAL INFORMATION:
APPLICANT: Sherr, Charles J.
APPLICANT: Quelle, Dawn E.
APPLICANT: Weber, Jason D.
APPLICANT: Rousset, Martine F.
APPLICANT: Frederique Zidny

TITLE OF INVENTION: ARF-p19, A NOVEL REGULATOR OF THE MAMMALIAN CELL CYCLE
FILE REFERENCE: 1340-1-023 CIP 1
CURRENT APPLICATION NUMBER: US/09/480,718
CURRENT FILING DATE: 2000-01-07
EARLIER APPLICATION NUMBER: 09/129,855
EARLIER FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 48
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4


```

; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-480-718-4

Query Match
Best Local Similarity 1.7%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVALVL 22
   |||||
Db 41 AAVALVL 47

RESULT 30
US-09-610-833-4
; Sequence 4, Application US/09610833
; Patent No. 6482929
; GENERAL INFORMATION:
; APPLICANT: Sherr, Charles, J.
; TITLE OF INVENTION: ARR-p19, A No. 6482929el Regulator of the
; Mammalian Cell Cycle
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/610,833
; FILING DATE: 06-Jul-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,470
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/534,975
; FILING DATE: 28-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel, L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0656.0590000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2500
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-610-833-4

Query Match
Best Local Similarity 1.7%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVALVL 22
   |||||
Db 41 AAVALVL 47

RESULT 31
US-09-129-855A-4
; Sequence 4, Application US/09129855A

; Patent No. 6586203
; GENERAL INFORMATION:
; APPLICANT: Sherr, Charles, J.
; TITLE OF INVENTION: ARR-p19, A No. 6586203el Regulator of the Mammalian
; Cell Cycle
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/129,855A
; FILING DATE: 06-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/954,470
; FILING DATE: 20-OCT-1997
; APPLICATION NUMBER: 08/534,975
; FILING DATE: 27-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-023 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-129-855A-4

Query Match
Best Local Similarity 1.7%; Score 7; DB 4; Length 132;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAVALVL 22
   |||||
Db 41 AAVALVL 47

RESULT 32
US-09-252-991A-18327
; Sequence 18327, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18327
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
```

US-09-252-991A-18327

Query Match 1.7%: Score 7; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ALAAVA 19
11111111
DB 101 ALAAVA 107

RESULT 33

US-08-145-995A-7
; Sequence 7, Application US/08145995A
; Patent No. 5482850
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTILDE K.S.
; APPLICANT: PAGE, ANTONY
; TITLE OF INVENTION: METHOD FOR IDENTIFYING ANTI-PARASITIC
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
; ADDRESS: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,995A
; FILING DATE: 29-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 43406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 169 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-145-995A-7

Query Match 1.7%: Score 7; DB 1; Length 169;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GQOPAGR 40
11111111
DB 11 GQOPAGR 17

RESULT 34
US-08-451-747-7
; Sequence 7, Application US/08451747
; Patent No. 5821107
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTILDE K.S.
; APPLICANT: PAGE, ANTONY
; TITLE OF INVENTION: METHOD FOR IDENTIFYING ANTI-PARASITIC
; TITLE OF INVENTION: COMPOUNDS

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREGORY D. WILLIAMS, NEW ENGLAND BIOLOGICALS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01915

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,747
FILING DATE:

CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/145,995
FILING DATE: 29-OCT-1993

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901

REFERENCE/DOCKET NUMBER: NEB-046-DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705

TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 169 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-451-747-7

Query Match 1.7%: Score 7; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GQOPAGR 40
11111111
DB 11 GQOPAGR 17

RESULT 35

US-09-134-852-7
; Sequence 7, Application US/09134852
; Patent No. 6127148
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTILDE K.S.
; APPLICANT: PAGE, ANTONY
; TITLE OF INVENTION: METHOD FOR IDENTIFYING ANTI-PARASITIC
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
; ADDRESS: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,852
; FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/145,995
FILING DATE: 29-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 43406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STR UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 169 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-134-852-7

Query Match 1.7%; Score 7; DB 3; Length 169;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GGOAPGR 40
|||
Db 11 GGOAPGR 17

RESULT 36
US-09-328-352-6419
Sequence 6419, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6419
LENGTH: 195
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6419

Query Match 1.7%; Score 7; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 AKRSAA 162
|||
Db 86 AKRSAA 92

RESULT 37
US-09-252-991A-25479
Sequence 25479, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 25479
LENGTH: 200
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25479

Query Match 1.7%; Score 7; DB 4; Length 200;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GGOAPGR 40
|||
Db 124 GGOAPGR 130

RESULT 38
US-09-252-991A-22833
Sequence 22833, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22833
LENGTH: 262
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22833

Query Match 1.7%; Score 7; DB 4; Length 262;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALAA 17
|||
Db 18 AAALAA 24

RESULT 39
US-09-107-532A-4013
Sequence 4013, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A. Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4013:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...275
SEQUENCE DESCRIPTION: SEQ ID NO: 4013:
US-09-107-532A-4013

Query Match 1.7%; Score 7; DB 4; Length 275;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 AVTRGAK 325
Db 70 AVTRGAK 76

RESULT 40
US-09-107-532A-5035
Sequence 5035, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5035:
SEQUENCE CHARACTERISTICS:
LENGTH: 287 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...287
SEQUENCE DESCRIPTION: SEQ ID NO: 5035:
US-09-107-532A-5035

Query Match 1.7%; Score 7; DB 4; Length 287;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 VKAAQAA 170
Db 92 VKAAQAA 98

Search completed: September 8, 2003, 14:11:00
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:02:47; Search time 45 Seconds

(without alignments)
1453.230 Million cell updates/sec

Title: US-09-889-756a-2

Perfect score: 412

Sequence: 1 MAFVAFKAMRAAALAAVAL.....AAPGSGVOTASEAKTASEAE 412

Scoring table: OLIGO

Gapop 60.0, Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: A_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412	100.0	412	21	AA190283
2	133	32.3	412	23	AAU91064
3	100	24.3	412	24	ABP78092
4	10	2.4	212	21	AA170472
5	10	2.4	222	22	AA193326
6	9	2.2	448	22	AA161543
7	9	2.2	19938	24	ABP76681
8	8	1.9	111	20	AA137704
9	8	1.9	171	22	AA198889

10	8	1.9	181	23	ABU05926	M. tuberculosis an
11	8	1.9	185	21	AA134400	Gene 30 human secr
12	8	1.9	194	21	AA186471	Human gene 51-enco
13	8	1.9	306	23	AB154244	Lactococcus lactis
14	8	1.9	318	23	AB154884	Bifidobacterium lo
15	8	1.9	426	22	AB128560	Novel human diagno
16	8	1.9	442	20	AA195347	Bifidobacterium lo
17	8	1.9	478	20	AA195347	Human methionine a
18	8	1.9	478	20	AA195347	Human p67 homology
19	8	1.9	478	20	AA195347	Mouse type 2 methi
20	8	1.9	478	20	AA195347	Rat type 2 methion
21	8	1.9	478	20	AA195347	Human type 2 methi
22	8	1.9	478	22	AB150275	Human type 2 methi
23	8	1.9	478	22	AB150275	Human type 2 methi
24	8	1.9	478	24	AB150275	Human type 2 methi
25	8	1.9	478	24	AB150275	Human type 2 methi
26	8	1.9	478	24	AB150275	Human type 2 methi
27	8	1.9	478	24	AB150275	Human type 2 methi
28	8	1.9	478	24	AB150275	Human type 2 methi
29	8	1.9	478	24	AB150275	Human type 2 methi
30	8	1.9	478	24	AB150275	Human type 2 methi
31	8	1.9	478	24	AB150275	Human type 2 methi
32	8	1.9	478	24	AB150275	Human type 2 methi
33	8	1.9	478	24	AB150275	Human type 2 methi
34	8	1.9	478	24	AB150275	Human type 2 methi
35	8	1.9	478	24	AB150275	Human type 2 methi
36	8	1.9	478	24	AB150275	Human type 2 methi
37	8	1.9	478	24	AB150275	Human type 2 methi
38	8	1.9	478	24	AB150275	Human type 2 methi
39	8	1.9	478	24	AB150275	Human type 2 methi
40	8	1.9	478	24	AB150275	Human type 2 methi
41	8	1.9	478	24	AB150275	Human type 2 methi
42	8	1.9	478	24	AB150275	Human type 2 methi
43	8	1.9	478	24	AB150275	Human type 2 methi
44	8	1.9	478	24	AB150275	Human type 2 methi
45	8	1.9	478	24	AB150275	Human type 2 methi
46	8	1.9	478	24	AB150275	Human type 2 methi
47	8	1.9	478	24	AB150275	Human type 2 methi
48	8	1.9	478	24	AB150275	Human type 2 methi
49	8	1.9	478	24	AB150275	Human type 2 methi
50	8	1.9	478	24	AB150275	Human type 2 methi
51	8	1.9	478	24	AB150275	Human type 2 methi
52	8	1.9	478	24	AB150275	Human type 2 methi
53	8	1.9	478	24	AB150275	Human type 2 methi
54	8	1.9	478	24	AB150275	Human type 2 methi
55	8	1.9	478	24	AB150275	Human type 2 methi
56	8	1.9	478	24	AB150275	Human type 2 methi
57	8	1.9	478	24	AB150275	Human type 2 methi
58	8	1.9	478	24	AB150275	Human type 2 methi
59	8	1.9	478	24	AB150275	Human type 2 methi
60	8	1.9	478	24	AB150275	Human type 2 methi
61	8	1.9	478	24	AB150275	Human type 2 methi
62	8	1.9	478	24	AB150275	Human type 2 methi
63	8	1.9	478	24	AB150275	Human type 2 methi
64	8	1.9	478	24	AB150275	Human type 2 methi
65	8	1.9	478	24	AB150275	Human type 2 methi
66	8	1.9	478	24	AB150275	Human type 2 methi
67	8	1.9	478	24	AB150275	Human type 2 methi
68	8	1.9	478	24	AB150275	Human type 2 methi
69	8	1.9	478	24	AB150275	Human type 2 methi
70	8	1.9	478	24	AB150275	Human type 2 methi
71	8	1.9	478	24	AB150275	Human type 2 methi
72	8	1.9	478	24	AB150275	Human type 2 methi
73	8	1.9	478	24	AB150275	Human type 2 methi
74	8	1.9	478	24	AB150275	Human type 2 methi
75	8	1.9	478	24	AB150275	Human type 2 methi
76	8	1.9	478	24	AB150275	Human type 2 methi
77	8	1.9	478	24	AB150275	Human type 2 methi
78	8	1.9	478	24	AB150275	Human type 2 methi
79	8	1.9	478	24	AB150275	Human type 2 methi
80	8	1.9	478	24	AB150275	Human type 2 methi
81	8	1.9	478	24	AB150275	Human type 2 methi
82	8	1.9	478	24	AB150275	Human type 2 methi

M. tuberculosis an
Gene 30 human secr
Human gene 51-enco
Lactococcus lactis
Bifidobacterium lo
Novel human diagno
Bifidobacterium lo
Human methionine a
Human p67 homology
Mouse type 2 methi
Rat type 2 methion
Human type 2 methi
eIF-2-associated p
Human methionine a
Human type 2 amino
Mouse type 2 amino
Human type 2 amino
Mouse type 2 amino
Protein differenti
Rat type 2 amino
Rat type 2 amino
Human colon cancer
Novel human diagno
Novel human diagno
Novel human diagno
E. coli cellular p
Salmonella typhi c
Drosophila melanog
Drosophila melanog
Novel human diagno
Human polypeptide
Human NOV103c prot
Human NOV103b prot
Human NOV103a prot
Human phosphodiester
Streptomyces virid
MHC binding peptid
Human macro protei
Mycobacterium tube
Mycobacterium tube
Human liver peptid
Peptide #3453 enco
Peptide #3485 enco
Protein #3388 enco
Human brain expres
Human bone marrow
Peptide #3424 enco
Peptide #3519 enco
Peptide #3379 enco
Human peptide enco
Fragment #3 of 7-a
Collagenase cleava
Human liver peptid
Peptide #7885 enco
Human brain expres
Human bone marrow
Peptide #8122 enco
Human peptide enco
Gene #40 human sec
Propionibacterium
Human musculoskele
Novel human muscu
Human novel protei
Human novel polype
Human ORFX protein
Novel human diagno
Antigen cc-26h. E
Human ORFX protein
Chlamydia pneumonia
Human 5' EST relat
Human ORFX ORF784
Human dehydrogenas
Synthetic elastome

83	7	1.7	114	22	AAW90521	Human immune/haema	156	7	1.7	300	23	ABP40481	Staphylococcus epi
84	7	1.7	114	22	AAW99922	Human polypeptide	157	7	1.7	302	21	AAW905003	Arabidopsis thalia
85	7	1.7	116	22	AAU49642	Propionibacterium	158	7	1.7	302	23	AAU93035	Arabidopsis trans
86	7	1.7	121	21	AAW04055	Human secreted pro	159	7	1.7	307	21	AAW33534	Arabidopsis thalia
87	7	1.7	122	21	AAW08770	Arabidopsis thalia	160	7	1.7	308	22	AAW90515	C glutamcum prote
88	7	1.7	122	21	AAW43679	Arabidopsis thalia	161	7	1.7	320	22	AAW91631	Arabidopsis thalia
89	7	1.7	123	22	AAU47983	Propionibacterium	162	7	1.7	332	21	AAW20285	Amino acid sequenc
90	7	1.7	123	22	AAU31822	Novel human secret	163	7	1.7	332	24	AAW99661	Arabidopsis thalia
91	7	1.7	125	21	AAW08769	Arabidopsis thalia	164	7	1.7	336	22	AAW99751	Oryza sativa perox
92	7	1.7	125	21	AAW43678	Arabidopsis thalia	165	7	1.7	339	22	AAU09964	Human cytidine dea
93	7	1.7	126	18	AAW06577	Arabidopsis thalia	166	7	1.7	360	21	AAW75619	Neisseria meningit
94	7	1.7	131	19	AAW42080	Lipo-oligosacchari	167	7	1.7	363	23	AAW93335	Herbicidally activ
95	7	1.7	132	18	AAW16323	The amino acid seq	168	7	1.7	373	24	ABP57036	Mycobacterium tube
96	7	1.7	132	21	AAV79145	Human ARF-p19, a n	169	7	1.7	373	24	ABP57037	Mycobacterium tube
97	7	1.7	132	22	AAW86208	Human ARF-p19 cell	170	7	1.7	374	12	AAW13987	Lycopene cyclase -
98	7	1.7	132	22	AAW25911	Human p1ARF tumor	171	7	1.7	374	17	AAW01125	lycopene cyclase.
99	7	1.7	132	23	ABP63705	Human ARF-p19 prot	172	7	1.7	374	18	AAW32474	Ewinnia heidicola
100	7	1.7	132	23	AAW15374	Human ORF75. Homo	173	7	1.7	392	21	AAW47099	Arabidopsis thalia
101	7	1.7	135	24	ABP76275	Human Arf protein.	174	7	1.7	393	24	ABP79495	N. gonorrhoeae ami
102	7	1.7	136	21	AAW44118	Human GENSER prote	175	7	1.7	394	21	AAW28531	Arabidopsis thalia
103	7	1.7	138	22	AAW41753	Arabidopsis thalia	176	7	1.7	394	23	AAW93276	C. albicans BAX-as
104	7	1.7	144	21	AAW42090	Human polypeptide	177	7	1.7	396	22	AAW81180	Mycobacterium tube
105	7	1.7	148	21	AAW01571	Arabidopsis thalia	178	7	1.7	400	21	AAW28530	Arabidopsis thalia
106	7	1.7	152	22	ABW69617	Human secreted pro	179	7	1.7	402	23	ABP39293	Staphylococcus epi
107	7	1.7	153	21	AAW57110	Drosophila melanog	180	7	1.7	411	21	AAW26371	Arabidopsis thalia
108	7	1.7	163	21	AAW47117	Human prostate can	181	7	1.7	418	19	AAW58603	Recombinant transc
109	7	1.7	166	22	AAW58218	Arabidopsis thalia	182	7	1.7	425	21	AAW38770	Arabidopsis thalia
110	7	1.7	173	21	AAW34362	Propionibacterium	183	7	1.7	426	21	AAW26370	Arabidopsis thalia
111	7	1.7	173	24	ABW48207	Arabidopsis thalia	184	7	1.7	429	21	AAW47098	Arabidopsis thalia
112	7	1.7	173	24	ABW56475	Human bladder canc	185	7	1.7	434	21	AAW58798	Breat and ovarian
113	7	1.7	176	21	AAW45174	Lung cancer-associ	186	7	1.7	439	21	AAW25538	Arabidopsis thalia
114	7	1.7	176	21	AAW45175	Gene 2 human secre	187	7	1.7	439	21	AAW43131	Arabidopsis thalia
115	7	1.7	179	22	ABW11529	Human secreted pro	188	7	1.7	444	22	ABW66980	Drosophila melanog
116	7	1.7	181	22	AAW41136	Human polypeptide	189	7	1.7	448	20	AAW74468	N. takasagoensis c
117	7	1.7	189	23	AAW18325	Keratin-associated	190	7	1.7	454	23	ABP43593	FJ10512. f18 clone
118	7	1.7	182	24	ABP78490	N. gonorrhoeae ami	191	7	1.7	461	23	AAW64816	Human albumin fusi
119	7	1.7	197	9	ABP80683	Phosphonitrilic r	192	7	1.7	461	23	AAW96176	Human secreted pro
120	7	1.7	200	21	AAW44117	Arabidopsis thalia	193	7	1.7	468	21	AAW20284	Arabidopsis thalia
121	7	1.7	200	21	AAW47116	Arabidopsis thalia	194	7	1.7	468	21	AAW54345	Arabidopsis thalia
122	7	1.7	203	21	AAW43068	Human ORF ORF2832	195	7	1.7	469	21	AAW31077	Arabidopsis thalia
123	7	1.7	203	22	ABW10250	Human cDNA SEQ ID	196	7	1.7	470	22	ABW60142	Drosophila melanog
124	7	1.7	203	22	AAW39350	Human polypeptide	197	7	1.7	471	21	AAW20283	Arabidopsis thalia
125	7	1.7	203	23	ABW66837	Human polypeptide	198	7	1.7	471	22	AAW79028	Human protein SEQ
126	7	1.7	203	23	AAW49382	Human macro protei	199	7	1.7	472	21	AAW54344	Arabidopsis thalia
127	7	1.7	205	9	AAW80916	Sequence of bovine	200	7	1.7	473	21	AAW31076	Arabidopsis thalia
128	7	1.7	206	22	AAW39238	Propionibacterium	201	7	1.7	477	22	ABW66931	Drosophila melanog
129	7	1.7	210	21	AAW34361	Arabidopsis thalia	202	7	1.7	479	21	AAW54343	Arabidopsis thalia
130	7	1.7	218	22	ABW62076	Drosophila melanog	203	7	1.7	480	21	AAW31075	Arabidopsis thalia
131	7	1.7	225	21	AAW45121	Human secreted pro	204	7	1.7	482	20	AAW95635	Homo sapiens secre
132	7	1.7	227	22	AAW57556	Propionibacterium	205	7	1.7	482	22	AAW80012	Human protein SEQ
133	7	1.7	233	21	AAW42089	Arabidopsis thalia	206	7	1.7	484	21	AAW28529	Arabidopsis thalia
134	7	1.7	236	22	AAW92898	C glutamcum prote	207	7	1.7	486	21	AAW47097	Arabidopsis thalia
135	7	1.7	252	22	AAW79289	Human protein SEQ	208	7	1.7	487	21	AAW33456	Human PRO1917 prot
136	7	1.7	253	21	AAW05005	Arabidopsis thalia	209	7	1.7	487	21	AAW96734	Human PRO1917 prot
137	7	1.7	255	21	AAW42088	Arabidopsis thalia	210	7	1.7	487	22	AAU12435	Human PRO1917 poly
138	7	1.7	257	21	AAW47115	Arabidopsis thalia	211	7	1.7	487	22	AAW81291	Human ARF protein
139	7	1.7	257	23	ABP03090	Human ORF protein	212	7	1.7	487	22	AAW50929	Human PRO1917 prot
140	7	1.7	259	21	AAW43535	Arabidopsis thalia	213	7	1.7	487	22	AAW64383	Amino acid sequenc
141	7	1.7	263	22	ABW16837	Novel human diapo	214	7	1.7	487	23	AAW31400	Human PRO1917 poly
142	7	1.7	267	21	AAW43460	Arabidopsis thalia	215	7	1.7	487	23	AAW91360	Novel human secret
143	7	1.7	269	21	AAW05087	Arabidopsis thalia	216	7	1.7	487	23	AAW923708	Human secreted pro
144	7	1.7	271	21	AAW77960	A. thaliana enviro	217	7	1.7	487	23	ABW95547	Human angiotensin
145	7	1.7	272	24	ABW78100	N. gonorrhoeae ami	218	7	1.7	487	23	ABW84941	Human PRO1917 prot
146	7	1.7	275	22	AAW35249	Enterococcus faeca	219	7	1.7	487	23	AAW81693	Human PRO protein,
147	7	1.7	275	23	AAW05677	M. tuberculosis an	220	7	1.7	487	24	ABW66833	Human PRO polypept
148	7	1.7	279	24	ABW70807	Human adipocyte se	221	7	1.7	487	24	ABW67109	Human secreted/lra
149	7	1.7	281	21	AAW05004	Arabidopsis thalia	222	7	1.7	487	24	ABW59914	Novel secreted and
150	7	1.7	281	22	AAW48050	Signal transductio	223	7	1.7	487	24	ABW60812	Human secreted/lra
151	7	1.7	288	21	AAW3810	Arabidopsis thalia	224	7	1.7	487	24	ABW73313	Human PRO1917 poly
152	7	1.7	290	21	AAW05086	Arabidopsis thalia	225	7	1.7	489	22	ABW30318	Novel human diapo
153	7	1.7	292	21	AAW06078	Arabidopsis thalia	226	7	1.7	492	20	AAW98121	Arabidopsis endo-1
154	7	1.7	295	21	AAW26372	Arabidopsis thalia	227	7	1.7	496	22	AAW47284	Strawberry cell.
155	7	1.7	295	23	ABW48857	Listeria monocycog	228	7	1.7	498	22	AAW929900	Human protein sequ

229	7	1.7	498	23	ABP43592	302	7	1.7	1078	23	ABB80736	Collagen type III-
230	7	1.7	500	24	ABP77032	303	7	1.7	1078	23	ABB09628	Amino acid sequenc
231	7	1.7	514	23	AAW48334	304	7	1.7	1078	23	AAE16478	Human collagen alp
232	7	1.7	515	23	ABP28186	305	7	1.7	1082	23	ABP62883	Human polypeptide
233	7	1.7	516	21	AAW25537	306	7	1.7	1090	22	AAW94737	Human protein sequ
234	7	1.7	516	21	AAW43130	307	7	1.7	1098	22	ABG03347	Novel human diagn
235	7	1.7	528	23	AAU93139	308	7	1.7	1144	21	AAW20007	Type III adenylLy1
236	7	1.7	528	23	AAU91399	309	7	1.7	1186	15	AAW56449	Delta endotoxin fir
237	7	1.7	536	23	AAU92991	310	7	1.7	1186	18	AAW1503	Nematode toxin 80f
238	7	1.7	538	22	ABW61067	311	7	1.7	1186	18	AAW21694	Bacillus thuringie
239	7	1.7	539	20	AAW23914	312	7	1.7	1186	18	AAW10652	Bacillus thuringie
240	7	1.7	539	23	AAU97744	313	7	1.7	1196	13	AAW28916	Type III procollag
241	7	1.7	539	23	AAU98656	314	7	1.7	1212	21	AAW33158	Human colorectal c
242	7	1.7	541	21	AAW38769	315	7	1.7	1212	22	AAW37061	Human breast cance
243	7	1.7	541	23	ABW91617	316	7	1.7	1257	22	ABW37062	Novel human diagn
244	7	1.7	542	21	AAW25536	317	7	1.7	1273	22	ABW00113	Novel human diagn
245	7	1.7	542	21	AAW43129	318	7	1.7	1277	22	AAW76536	Corynebacterium gl
246	7	1.7	549	22	ABW13328	319	7	1.7	1353	22	AAW79520	Human protein SBQ
247	7	1.7	549	22	ABW25686	320	7	1.7	1367	22	AAW78536	Human protein SBQ
248	7	1.7	555	21	AAW38768	321	7	1.7	1466	22	ABW50291	Collagen type III
249	7	1.7	556	23	ABP73668	322	7	1.7	1466	23	ABW90747	Human Tumour Endot
250	7	1.7	559	22	AAW78707	323	7	1.7	1466	24	ABW47418	Breast cancer asso
251	7	1.7	561	22	ABW27440	324	7	1.7	1466	24	ABW54454	Human tumour endot
252	7	1.7	568	22	ABW28701	325	7	1.7	1469	22	ABW15191	Novel human diagn
253	7	1.7	570	23	AAW71312	326	7	1.7	1787	23	ABW49791	Listeria monocycog
254	7	1.7	570	23	AAU71858	327	7	1.7	1846	22	ABW63163	Drosophila melanog
255	7	1.7	572	19	AAW54159	328	7	1.7	2012	22	AAW76534	Corynebacterium gl
256	7	1.7	572	22	ABW82871	329	7	1.7	2139	22	ABW62393	Drosophila melanog
257	7	1.7	576	22	ABW12306	330	7	1.7	2993	22	ABW92485	C glutamyl prote
258	7	1.7	579	22	ABW29084	331	7	1.7	3079	22	AAW59926	GAP protein Itaz.
259	7	1.7	591	21	AAW27325	332	7	1.7	4132	22	ABW99870	S. cinnamomensis M
260	7	1.7	599	22	ABW24227	333	7	1.7	7068	22	AAW10142	Streptomyces nous
261	7	1.7	600	23	ABW65869	334	7	1.7	9477	22	ABW10144	Streptomyces nous
262	7	1.7	602	23	ABW68917	335	7	1.7	9477	22	ABW47131	Human Blvs binding
263	7	1.7	615	21	ABW24068	336	7	1.5	8	19	AAW40855	Cytotoxic Epstein-
264	7	1.7	615	21	AAW66678	337	7	1.5	9	20	AAW10577	HLA Class I motif
265	7	1.7	615	22	AAW65201	338	7	1.5	9	20	AAW10193	T cell epitope/MHC
266	7	1.7	615	22	ABW59094	339	7	1.5	9	22	ABW12579	HIV A02 super moti
267	7	1.7	615	24	ABW59241	340	7	1.5	9	22	ABW12580	HIV A02 super moti
268	7	1.7	615	24	ABW59390	341	7	1.5	9	22	ABW20808	HIV A03 motif gag
269	7	1.7	615	24	ABW60525	342	7	1.5	9	22	ABW20809	HIV A03 motif gag
270	7	1.7	615	24	ABW58016	343	7	1.5	9	23	ABW79875	MHC class I molecu
271	7	1.7	615	24	ABW58947	344	7	1.5	9	23	ABW680260	MHC class I molecu
272	7	1.7	615	24	ABU13907	345	7	1.5	9	24	ABW96478	HLA class I molecu
273	7	1.7	615	24	ABU10862	346	7	1.5	9	24	ABW06155	Human cancer-relat
274	7	1.7	622	23	ABW91788	347	7	1.5	9	24	ABW06466	Human cancer-relat
275	7	1.7	631	21	AAW59208	348	7	1.5	9	24	ABW06675	Human cancer-relat
276	7	1.7	631	24	ABW97686	349	7	1.5	10	7	AAW60730	Block unit. Synth
277	7	1.7	631	24	AAO26962	350	7	1.5	10	7	ABW15598	HIV A24 super moti
278	7	1.7	644	22	AAW70158	351	7	1.5	10	22	ABW15599	HIV A24 super moti
279	7	1.7	645	22	ABW61545	352	7	1.5	10	22	ABW18757	HIV B62 super moti
280	7	1.7	662	22	ABW59430	353	7	1.5	10	22	ABW18899	HIV B62 super moti
281	7	1.7	665	15	AAW47189	354	7	1.5	10	22	ABW20452	HIV A03 motif gag
282	7	1.7	674	22	AAW93773	355	7	1.5	10	22	ABW20810	HIV A03 motif gag
283	7	1.7	674	22	AAW92883	356	7	1.5	10	22	AAW95637	Human complementat
284	7	1.7	687	24	ABW77385	357	7	1.5	10	23	AAW18577	Area-specific modt
285	7	1.7	691	22	AAW58013	358	7	1.5	10	24	ABW06837	Human cancer-relat
286	7	1.7	735	23	AAO21849	359	7	1.5	11	11	AAW08088	Artifreeze segment
287	7	1.7	738	19	AAW56163	360	7	1.5	11	20	AAW95680	Pneumocystis carin
288	7	1.7	745	23	ABW65572	361	7	1.5	11	21	AAW21136	Sic homolog 3 dom
289	7	1.7	749	24	ABW41749	362	7	1.5	11	21	ABW12811	HIV A02 super moti
290	7	1.7	848	17	AAW43716	363	7	1.5	11	22	ABW12812	HIV A24 super moti
291	7	1.7	896	17	AAW00733	364	7	1.5	11	22	ABW15660	HIV A24 super moti
292	7	1.7	900	22	ABW61260	365	7	1.5	11	22	ABW15706	HIV A24 super moti
293	7	1.7	908	22	AAW81162	366	7	1.5	11	22	ABW15707	HIV A24 super moti
294	7	1.7	908	21	AAW54055	367	7	1.5	11	22	ABW17948	HIV B58 super moti
295	7	1.7	912	21	AAW54057	368	7	1.5	11	22	ABW18761	HIV B62 super moti
296	7	1.7	917	13	AAW56334	369	7	1.5	11	22	ABW20465	HIV A03 motif gag
297	7	1.7	917	21	AAW55073	370	7	1.5	11	22	ABW20788	HIV A03 motif gag
298	7	1.7	970	23	ABW65559	371	7	1.5	11	22	ABW20811	HIV A03 motif gag
299	7	1.7	1036	23	ABW91801	372	7	1.5	11	22	ABW20812	HIV A03 motif gag
300	7	1.7	1078	16	AAW71704	373	7	1.5	12	20	AAW95681	Pneumocystis carin
301	7	1.7	1078	21	AAW96125	374	7	1.5	12	22	ABW12909	HIV A02 super moti

375	6	1.5	12	22	ABP12910	HIV A02 super moti	448	6	1.5	27	17	AAR98018	Fusogenic peptide
376	6	1.5	12	22	ABP12911	HIV A02 super moti	449	6	1.5	27	22	AAB55313	Anti-HPV3 Fl regio
377	6	1.5	12	22	ABP12912	HIV A02 super moti	450	6	1.5	27	22	AAB55344	Anti-HPV3 Fl regio
378	6	1.5	12	22	ABP18746	HIV B62 super moti	451	6	1.5	28	16	AAR76556	HPF3 peptide deriv
379	6	1.5	12	22	ABP18900	HIV B62 super moti	452	6	1.5	28	16	AAR64714	HPF3 peptide deriv
380	6	1.5	13	16	AAR87410	Human MHC class II	453	6	1.5	28	21	AAV88686	Core polypeptide f
381	6	1.5	13	21	AAB08007	Peptide which is w	454	6	1.5	28	21	AAV88687	Core polypeptide f
382	6	1.5	13	21	AAV99239	HLA class II bindi	455	6	1.5	28	21	AAV88688	Core polypeptide f
383	6	1.5	13	23	AAO18843	HIV Nef protein de	456	6	1.5	28	22	ABB00045	Viral DP178/107-1
384	6	1.5	13	24	ABP70245	Peptide YKL, speci	457	6	1.5	28	22	ABB00046	Viral DP178/107-1
385	6	1.5	13	24	ABP70245	Peptide YKL, speci	458	6	1.5	28	22	ABB00047	Viral DP178/107-1
386	6	1.5	14	14	AAR32963	Mastoparan analogu	459	6	1.5	28	22	ABB01509	Viral core polypep
387	6	1.5	15	24	ABR31652	Human cancer-relat	460	6	1.5	28	22	ABB01510	Viral core polypep
388	6	1.5	15	24	ABR31653	Human cancer-relat	461	6	1.5	28	22	ABB01511	Viral core polypep
389	6	1.5	15	24	ABR31670	Human cancer-relat	462	6	1.5	28	22	AAU12594	DP178-1like/DP107-1
390	6	1.5	15	24	ABR31671	Human cancer-relat	463	6	1.5	28	22	AAU12595	DP178-1like/DP107-1
391	6	1.5	15	24	ABR32182	Human cancer-relat	464	6	1.5	28	22	AAU12596	Core polypeptide T
392	6	1.5	15	24	ABR32185	Human cancer-relat	465	6	1.5	28	22	AAU12597	Core polypeptide T
393	6	1.5	15	24	ABR32183	Human cancer-relat	466	6	1.5	28	22	AAU12598	Core polypeptide T
394	6	1.5	15	24	ABR32184	Human cancer-relat	467	6	1.5	28	22	AAU12599	Core polypeptide T
395	6	1.5	15	24	ABR32277	Human cancer-relat	468	6	1.5	28	22	AAU12600	Core polypeptide T
396	6	1.5	15	24	ABR32277	Human cancer-relat	469	6	1.5	28	22	AAU12601	Core polypeptide T
397	6	1.5	16	20	AAV32997	Rat SSeCKS consens	470	6	1.5	29	16	AAR76557	HPF3 peptide deriv
398	6	1.5	17	16	AAR64703	Membrane forming a	471	6	1.5	29	16	AAR76558	HPF3 peptide deriv
399	6	1.5	17	22	AAB55303	Anti-HPV3 Fl regio	472	6	1.5	29	16	AAR76559	HPF3 peptide deriv
400	6	1.5	18	16	AAR64704	HPF3 peptide deriv	473	6	1.5	29	16	AAR76560	HPF3 peptide deriv
401	6	1.5	18	22	AAU04221	Chaperonin binding	474	6	1.5	29	22	AAB55315	Novel human diagno
402	6	1.5	18	22	AAB89367	HIV gp120 protein	475	6	1.5	29	22	AAB55316	Novel human diagno
403	6	1.5	18	22	AAB55304	Anti-HPV3 Fl regio	476	6	1.5	29	23	ABR76558	Anti-HPV3 Fl regio
404	6	1.5	19	22	AAB28627	Human K+betaM2 pro	477	6	1.5	30	16	AAR64716	HPF3 peptide deriv
405	6	1.5	19	22	AAR64705	HPF3 peptide deriv	478	6	1.5	30	16	AAR64717	HPF3 peptide deriv
406	6	1.5	20	16	AAB55305	Anti-HPV3 Fl regio	479	6	1.5	30	22	ABG13834	Novel human diagno
407	6	1.5	20	16	AAR64706	HPF3 peptide deriv	480	6	1.5	30	22	ABG13835	Novel human diagno
408	6	1.5	20	17	AAR89341	Cdk2-interacting p	481	6	1.5	30	22	AAB55316	Anti-HPV3 Fl regio
409	6	1.5	20	18	AAW32116	Interaction trap s	482	6	1.5	30	22	AAB55347	Anti-HPV3 Fl regio
410	6	1.5	20	21	AAV49332	Peptide Seq ID No:	483	6	1.5	31	16	AAR76559	HPF3 peptide deriv
411	6	1.5	21	16	AAR64707	Anti-HPV3 Fl regio	484	6	1.5	31	16	AAR64717	HPF3 peptide deriv
412	6	1.5	21	22	AAB55307	HPF3 peptide deriv	485	6	1.5	31	22	AAB55317	Anti-HPV3 Fl regio
413	6	1.5	21	22	AAB55308	Anti-HPV3 Fl regio	486	6	1.5	31	22	AAB55318	Anti-HPV3 Fl regio
414	6	1.5	22	16	AAR64708	HPF3 peptide deriv	487	6	1.5	31	22	AAR64719	HPF3 peptide deriv
415	6	1.5	22	20	AAV19768	SEQ ID NO 486 from	488	6	1.5	31	22	AAR64720	HPF3 peptide deriv
416	6	1.5	22	21	AAB51604	Yada homologous pe	489	6	1.5	32	16	AAR64721	HPF3 peptide deriv
417	6	1.5	22	22	AAB55308	Anti-HPV3 Fl regio	490	6	1.5	32	16	AAR64722	HPF3 peptide deriv
418	6	1.5	23	16	AAR64709	HPF3 peptide deriv	491	6	1.5	32	22	AAB55319	Anti-HPV3 Fl regio
419	6	1.5	23	20	AAW92516	Human FEN-1 protei	492	6	1.5	32	22	AAB55320	Anti-HPV3 Fl regio
420	6	1.5	23	22	AAB55309	Human DNA polymera	493	6	1.5	32	22	AAB55321	Human DNA polymera
421	6	1.5	23	23	AAB20093	Anti-HPV3 Fl regio	494	6	1.5	33	16	AAR76561	HPF3 peptide deriv
422	6	1.5	23	23	AAB04556	Hepatitis C capsid	495	6	1.5	33	16	AAR76562	HPF3 peptide deriv
423	6	1.5	24	16	AAR76552	HPF3 peptide deriv	496	6	1.5	33	22	ABG1428	Peptide #4147 enco
424	6	1.5	24	16	AAR64710	HPF3 peptide deriv	497	6	1.5	33	22	ABG1429	Peptide #4147 enco
425	6	1.5	24	17	AAR89353	Medium binding pep	498	6	1.5	33	22	ABG1430	Peptide #4147 enco
426	6	1.5	24	21	AAB51532	Proteobacterial ex	499	6	1.5	33	22	ABG1431	Peptide #4147 enco
427	6	1.5	24	21	AAV49335	Cdk2 interacting m	500	6	1.5	33	22	AAV49336	Human bone marrow
428	6	1.5	24	22	AAB55310	Anti-HPV3 Fl regio	501	6	1.5	33	22	AAV49337	Human bone marrow
429	6	1.5	25	16	AAR76553	Anti-HPV3 Fl regio	502	6	1.5	33	22	AAV49338	Human bone marrow
430	6	1.5	25	16	AAR76553	HPF3 peptide deriv	503	6	1.5	33	22	AAV49339	Human bone marrow
431	6	1.5	25	21	AAV96541	HPF3 peptide deriv	504	6	1.5	33	22	AAV96542	Human bone marrow
432	6	1.5	25	21	AAV96541	Mutant human NF-AT	505	6	1.5	33	22	AAV96543	Human bone marrow
433	6	1.5	25	22	AAB55311	Human AP-ATC1 SRR	506	6	1.5	33	22	AAB55312	Human bone marrow
434	6	1.5	25	22	AAB55311	Anti-HPV3 Fl regio	507	6	1.5	33	22	AAB55313	Human bone marrow
435	6	1.5	25	22	AAB55311	Anti-HPV3 Fl regio	508	6	1.5	33	22	AAB55314	Human bone marrow
436	6	1.5	25	23	ABG62524	Eubacterial MutS1	509	6	1.5	34	16	AAR64723	HPF3 peptide deriv
437	6	1.5	25	23	ABG62525	Eubacterial MutS1	510	6	1.5	34	16	AAR64724	HPF3 peptide deriv
438	6	1.5	25	23	ABG62611	Eubacterial MutS1	511	6	1.5	34	16	AAR64725	HPF3 peptide deriv
439	6	1.5	25	23	AAE20992	Human DNA polymera	512	6	1.5	34	16	AAE20993	Human DNA polymera
440	6	1.5	25	23	AAE20992	Streptococcus pyog	513	6	1.5	34	16	AAE20994	Streptococcus pyog
441	6	1.5	26	16	AAR64712	HPF3 peptide deriv	514	6	1.5	35	6	AAV50187	HPF3 peptide deriv
442	6	1.5	26	16	AAR64712	HPF3 peptide deriv	515	6	1.5	35	6	AAV50188	HPF3 peptide deriv
443	6	1.5	26	22	AAV99892	Human excretory re	516	6	1.5	35	20	AAR64726	Core polypeptide f
444	6	1.5	26	22	AAV99892	Human bladder anti	517	6	1.5	35	20	AAV99893	Core polypeptide f
445	6	1.5	26	22	AAB55312	Anti-HPV3 Fl regio	518	6	1.5	35	21	AAV88689	Core polypeptide f
446	6	1.5	27	16	AAR76555	HPF3 peptide deriv	519	6	1.5	35	21	AAV88690	Core polypeptide f
447	6	1.5	27	16	AAR76555	HPF3 peptide deriv	520	6	1.5	35	21	AAV88691	Core polypeptide f

521	6	1.5	35	21	AAV88805	Core polypeptide f	594	6	1.5	35	22	AA877168	Core polypeptide T
522	6	1.5	35	21	AAV88806	Core polypeptide f	595	6	1.5	35	22	AA854832	HPIV antiviral act
523	6	1.5	35	21	AAV88807	Core polypeptide f	596	6	1.5	35	22	AA855397	Anti-HPV3 peptide
524	6	1.5	35	21	AAV88808	Core polypeptide f	597	6	1.5	36	22	AB858245	Human liver peptid
525	6	1.5	35	21	AAV88809	Core polypeptide f	598	6	1.5	36	22	AB842839	Peptide #10345 enc
526	6	1.5	35	21	AAV88810	Core polypeptide f	599	6	1.5	36	22	AA863731	Human brain expres
527	6	1.5	35	21	AAV88811	Core polypeptide f	600	6	1.5	36	22	AA876545	Human bone marrow
528	6	1.5	35	21	AAV88812	Core polypeptide f	601	6	1.5	36	22	AA836652	Peptide #10689 enc
529	6	1.5	35	21	AAV88813	Core polypeptide f	602	6	1.5	36	22	ABG45798	Human peptid enc
530	6	1.5	35	22	AB800043	Viral DP178/107-11	603	6	1.5	37	21	AA858871	Arabidopsis thalia
531	6	1.5	35	22	AB800161	Viral DP178/107-11	604	6	1.5	37	22	AA012442	Human polypeptide
532	6	1.5	35	22	AB800162	Viral DP178/107-11	605	6	1.5	37	22	AA804214	Human gene 17 enc
533	6	1.5	35	22	AB800163	Viral DP178/107-11	606	6	1.5	37	22	AA804238	Human gene 17 enc
534	6	1.5	35	22	AB800164	Viral DP178/107-11	607	6	1.5	37	23	AB8882209	Chimeric pPHIR ago
535	6	1.5	35	22	AB800165	Viral DP178/107-11	608	6	1.5	37	23	AB864512	Human albumin fusi
536	6	1.5	35	22	AB800166	Viral DP178/107-11	609	6	1.5	37	23	AB864513	Human albumin fusi
537	6	1.5	35	22	AB800167	Viral DP178/107-11	610	6	1.5	38	22	AB859559	Human liver peptid
538	6	1.5	35	22	AB800168	Viral DP178/107-11	611	6	1.5	38	22	AB844156	Peptide #11702 enc
539	6	1.5	35	22	AB800169	Viral DP178/107-11	612	6	1.5	38	22	AB827074	Protein #9073 enc
540	6	1.5	35	22	AB800170	Viral DP178/107-11	613	6	1.5	38	22	AA865227	Human brain expres
541	6	1.5	35	22	AB800171	Viral DP178/107-11	614	6	1.5	38	22	AA877930	Human bone marrow
542	6	1.5	35	22	AB800172	Viral DP178/107-11	615	6	1.5	38	22	AA821825	Peptide #8259 enc
543	6	1.5	35	22	AB801507	Viral core polypep	616	6	1.5	38	22	AA838149	Peptide #12186 enc
544	6	1.5	35	22	AB801626	Viral core polypep	617	6	1.5	38	23	AB846952	Human peptid enc
545	6	1.5	35	22	AB801627	Viral core polypep	618	6	1.5	39	22	AB802667	Novel human diagn
546	6	1.5	35	22	AB801628	Viral core polypep	619	6	1.5	39	22	AA822276	Human cardiovascu
547	6	1.5	35	22	AB801629	Viral core polypep	620	6	1.5	39	22	AA867535	Human acid sequenc
548	6	1.5	35	22	AB801630	Viral core polypep	621	6	1.5	39	24	AB899877	Breast specific po
549	6	1.5	35	22	AB801631	Viral core polypep	622	6	1.5	40	21	AA809856	Arabidopsis thalia
550	6	1.5	35	22	AB801632	Viral core polypep	623	6	1.5	40	22	AB848658	Human liver peptid
551	6	1.5	35	22	AB801633	Viral core polypep	624	6	1.5	40	22	AB828639	Peptide #1290 enc
552	6	1.5	35	22	AB801634	Viral core polypep	625	6	1.5	40	22	AB833823	Peptide #1262 enc
553	6	1.5	35	22	AB801635	Viral core polypep	626	6	1.5	40	22	AA819265	Protein #1264 enc
554	6	1.5	35	22	AB801636	Viral core polypep	627	6	1.5	40	22	AA854589	Human brain expres
555	6	1.5	35	22	AB801637	Viral core polypep	628	6	1.5	40	22	AA866995	Human bone marrow
556	6	1.5	35	22	AAU12582	DP178-1like/DP107-1	629	6	1.5	40	22	AA814855	Peptide #1289 enc
557	6	1.5	35	22	AAU12710	DP178-1like/DP107-1	630	6	1.5	40	22	AA827282	Peptide #1319 enc
558	6	1.5	35	22	AAU12711	DP178-1like/DP107-1	631	6	1.5	40	22	AA802580	Peptide #1262 enc
559	6	1.5	35	22	AAU12712	DP178-1like/DP107-1	632	6	1.5	40	22	AB836652	Human peptid enc
560	6	1.5	35	22	AAU12713	DP178-1like/DP107-1	633	6	1.5	41	22	AB857857	Human liver peptid
561	6	1.5	35	22	AAU12714	DP178-1like/DP107-1	634	6	1.5	41	22	AB842435	Peptide #9941 enc
562	6	1.5	35	22	AAU12715	DP178-1like/DP107-1	635	6	1.5	41	22	AB825885	Protein #7884 enc
563	6	1.5	35	22	AAU12716	DP178-1like/DP107-1	636	6	1.5	41	22	AA863321	Human brain expres
564	6	1.5	35	22	AAU12717	DP178-1like/DP107-1	637	6	1.5	41	22	AA876136	Human bone marrow
565	6	1.5	35	22	AAU12718	DP178-1like/DP107-1	638	6	1.5	41	22	AA836246	Peptide #10283 enc
566	6	1.5	35	22	AAU12719	DP178-1like/DP107-1	639	6	1.5	41	23	AB845502	Human peptid enc
567	6	1.5	35	22	AAU12720	DP178-1like/DP107-1	640	6	1.5	42	20	AA8700315	Human secreted pro
568	6	1.5	35	22	AAU12721	DP178-1like/DP107-1	641	6	1.5	43	22	AA804607	Human polypeptide
569	6	1.5	35	22	AAU14037	DP107-1like peptid	642	6	1.5	43	22	AA806106	Human polypeptide
570	6	1.5	35	22	AAU14038	DP107-1like peptid	643	6	1.5	43	22	AA862381	Human gene 20-enco
571	6	1.5	35	22	AAU14039	DP107-1like peptid	644	6	1.5	44	21	AA895964	Construct used in
572	6	1.5	35	22	AAU14040	DP107-1like peptid	645	6	1.5	44	21	AA809855	Arabidopsis thalia
573	6	1.5	35	22	AAU14041	DP107-1like peptid	646	6	1.5	44	21	AA833064	F-box motif of FBP
574	6	1.5	35	22	AAU14042	DP107-1like peptid	647	6	1.5	44	21	AA833064	Human F-box motif
575	6	1.5	35	22	AAU14043	DP107-1like peptid	648	6	1.5	44	21	AA82488	Gene 29 human secr
576	6	1.5	35	22	AAU14044	DP107-1like peptid	649	6	1.5	46	21	AA851814	Novel human diagn
577	6	1.5	35	22	AAU14045	DP107-1like peptid	650	6	1.5	46	22	AB827732	Novel human diagn
578	6	1.5	35	22	AAU14046	DP107-1like peptid	651	6	1.5	46	22	AA804462	Human polypeptide
579	6	1.5	35	22	AAU14047	DP107-1like peptid	652	6	1.5	47	23	AB825481	Streptococcus poly
580	6	1.5	35	22	AAU14048	DP107-1like peptid	653	6	1.5	48	22	AB824708	Novel human diagn
581	6	1.5	35	22	AA892291	Virus related pept	654	6	1.5	49	22	AA858330	Human brain expres
582	6	1.5	35	22	AA877039	Core polypeptide T	655	6	1.5	49	22	AA870860	Human bone marrow
583	6	1.5	35	22	AA877157	Core polypeptide T	656	6	1.5	49	22	AA804481	Human polypeptide
584	6	1.5	35	22	AA877158	Core polypeptide T	657	6	1.5	49	22	AA818647	Peptide #5081 enc
585	6	1.5	35	22	AA877159	Core polypeptide T	658	6	1.5	50	22	AB816627	Human nervous syst
586	6	1.5	35	22	AA877160	Core polypeptide T	659	6	1.5	51	21	AA834355	Human secreted pro
587	6	1.5	35	22	AA877161	Core polypeptide T	660	6	1.5	51	22	AA8007406	Human polypeptide
588	6	1.5	35	22	AA877162	Core polypeptide T	661	6	1.5	52	22	AB853747	Human liver peptid
589	6	1.5	35	22	AA877163	Core polypeptide T	662	6	1.5	52	22	AA856076	Propionibacterium
590	6	1.5	35	22	AA877164	Core polypeptide T	663	6	1.5	52	22	AB838853	Peptide #6359 enc
591	6	1.5	35	22	AA877165	Core polypeptide T	664	6	1.5	52	22	AA859498	Human brain expres
592	6	1.5	35	22	AA877166	Core polypeptide T	665	6	1.5	52	22	AA872063	Human bone marrow
593	6	1.5	35	22	AA877167	Core polypeptide T	666	6	1.5	52	22	AA885536	Human immune/haema

667	6	1.5	52	22	AA032328	Peptide #6365 enco	740	6	1.5	67	19	AA061094	Rat endogenous A6L
668	6	1.5	52	23	ABG41878	Human ORFX protein	741	6	1.5	67	22	AB030926	Human musculoskele
669	6	1.5	53	23	ABP03591	Human ORFX protein	742	6	1.5	67	23	AAE20986	Human DNA polymera
670	6	1.5	53	21	AA026631	Human secreted pro	743	6	1.5	67	24	ABU13220	Novel human muscul
671	6	1.5	53	21	AA01214	Human secreted pro	744	6	1.5	67	24	ABP06055	N. gonorrhoeae ami
672	6	1.5	53	22	AAU49365	Proionibacterium	745	6	1.5	68	19	AAW79441	Staphylococcus aur
673	6	1.5	53	22	ABBI0593	Human pancreatic c	746	6	1.5	68	19	AA061093	Human endogenous A
674	6	1.5	53	22	AA092346	Human digestive sy	747	6	1.5	68	21	AA059221	Human polypeptide
675	6	1.5	53	23	ABP02094	Human ORFX protein	748	6	1.5	68	22	AA012124	Human polypeptide
676	6	1.5	54	22	ABG52592	Human liver peptid	749	6	1.5	68	23	AAE20987	Human DNA polymera
677	6	1.5	54	22	AAU50583	Proionibacterium	750	6	1.5	68	24	ABR48066	Human secreted pro
678	6	1.5	54	22	ABP37777	Peptide #5283 enco	751	6	1.5	69	8	AAV70041	Secretory signal s
679	6	1.5	54	22	AB023061	Protein #5060 enco	752	6	1.5	69	21	AA059220	Arabidopsis thalia
680	6	1.5	54	22	AAU31671	Novel human secret	753	6	1.5	69	21	AA009365	Human polypeptide
681	6	1.5	54	22	AAU31671	Human brain expres	754	6	1.5	69	22	AA012636	Human polypeptide
682	6	1.5	54	22	AA070876	Human bone marrow	755	6	1.5	69	22	AA02640	Human protein sequ
683	6	1.5	54	22	AA012262	Human polypeptide	756	6	1.5	69	23	ABP34734	Human ORF3707 prot
684	6	1.5	54	22	AA012262	Peptide #5146 enco	757	6	1.5	70	14	AA039693	rpP4-307. Synthet
685	6	1.5	54	22	AA012262	Peptide #5205 enco	758	6	1.5	70	21	AA018679	Mutant peptide der
686	6	1.5	54	22	AA012262	Peptide #5146 enco	759	6	1.5	70	21	AA018679	Arabidopsis thalia
687	6	1.5	54	22	AA012262	Human peptide enco	760	6	1.5	70	21	AA018679	Novel human diagn
688	6	1.5	54	22	AA012262	Human haematologic	761	6	1.5	70	21	AA018679	Human haematologic
689	6	1.5	55	22	AA011362	Human haematologic	762	6	1.5	70	22	AA025188	DP107-like peptide
690	6	1.5	55	22	AA011362	Human acid sequenc	763	6	1.5	70	22	AA025188	Virus related pept
691	6	1.5	55	22	AA011362	Arabidopsis thalia	764	6	1.5	70	22	AA025188	HIV antiviral act
692	6	1.5	56	22	ABG53716	Human liver peptid	765	6	1.5	70	22	AA025188	Human ovary specif
693	6	1.5	56	22	ABU48344	Proionibacterium	766	6	1.5	70	22	AB030371	Human secreted pro
694	6	1.5	56	22	ABU48344	Peptide #6330 enco	767	6	1.5	71	21	AB030371	Human testicular a
695	6	1.5	56	22	ABU48344	Human nervous syst	768	6	1.5	71	21	AB030371	Proionibacterium
696	6	1.5	56	22	ABU48344	Protein #5844 enco	769	6	1.5	71	22	AA025188	Novel human secret
697	6	1.5	56	22	AA025188	Human bone marrow	770	6	1.5	71	22	AA025188	Human reproductive
698	6	1.5	56	22	AA025188	Human brain expres	771	6	1.5	71	22	AA025188	Human DTHP enzyme
699	6	1.5	56	22	AA025188	Human immune/haema	772	6	1.5	71	22	AA025188	Proionibacterium
700	6	1.5	56	22	AA025188	Peptide #6335 enco	773	6	1.5	71	22	AA025188	Novel human diagn
701	6	1.5	56	22	AA025188	Human peptide enco	774	6	1.5	71	22	AA025188	Novel human diagn
702	6	1.5	56	22	AA025188	Human ORFX protein	775	6	1.5	72	22	AA025188	Human nervous syst
703	6	1.5	56	22	AA025188	Human novel foetal	776	6	1.5	72	22	AA025188	Human polypeptide
704	6	1.5	57	23	AAE20982	Human DNA polymera	777	6	1.5	72	22	AA025188	Human DTHP recept
705	6	1.5	57	23	ABP26366	Streptococcus poly	778	6	1.5	72	22	AA025188	Novel human secret
706	6	1.5	58	22	AAU36171	Klebsiella pneumon	779	6	1.5	72	22	AA025188	Human nervous syst
707	6	1.5	58	22	AA000928	Human polypeptide	780	6	1.5	72	22	AA025188	Human secreted pro
708	6	1.5	58	22	AA000928	Human polypeptide	781	6	1.5	72	22	AA025188	Arabidopsis thalia
709	6	1.5	58	22	AA000928	Human acid sequenc	782	6	1.5	73	21	AA025188	Human secreted pro
710	6	1.5	58	22	AA000928	Human DNA polymera	783	6	1.5	73	21	AA025188	C glutamicum prote
711	6	1.5	58	22	AA000928	Human acid sequenc	784	6	1.5	73	22	AA025188	Arabidopsis thalia
712	6	1.5	59	19	AAW63840	Human ORFX protein	785	6	1.5	74	21	AA025188	Proionibacterium
713	6	1.5	59	19	AAW63840	H. pylori strain J	786	6	1.5	74	21	AA025188	Human nervous syst
714	6	1.5	59	20	AAW63840	Helicobacter pylor	787	6	1.5	74	22	AA025188	Novel human secret
715	6	1.5	59	20	AAW63840	Helicobacter pylor	788	6	1.5	74	22	AA025188	Human polypeptide
716	6	1.5	59	20	AAW63840	Helicobacter pylor	789	6	1.5	74	22	AA025188	E. coli strain O15
717	6	1.5	59	20	AAW63840	Human pancreatic c	790	6	1.5	74	22	AA025188	Toxoplasma gondii
718	6	1.5	59	22	AAW63840	Human digestive sy	791	6	1.5	75	12	AA025188	Human secreted pro
719	6	1.5	60	9	AA0634	Sequence encoded	792	6	1.5	75	20	AA025188	Human DTHP recept
720	6	1.5	60	22	AAU65254	Proionibacterium	793	6	1.5	76	22	AA025188	Human nervous syst
721	6	1.5	60	22	AAU65254	Proionibacterium	794	6	1.5	76	22	AA025188	Human nervous syst
722	6	1.5	61	20	AA088545	Secreted protein e	795	6	1.5	76	22	AA025188	Human nervous syst
723	6	1.5	61	22	AAU52600	Proionibacterium	796	6	1.5	77	20	AA025188	Human secreted pro
724	6	1.5	61	22	AAU52600	Human secreted pro	797	6	1.5	77	21	AA025188	Novel human secret
725	6	1.5	61	22	AAU52600	Human polypeptide	798	6	1.5	77	22	AA025188	Human rhinovirus t
726	6	1.5	61	22	AAU52600	Human polypeptide	799	6	1.5	77	22	AA025188	Human liver peptid
727	6	1.5	62	21	AAU52600	Human ORFX protein	800	6	1.5	78	22	AA025188	Proionibacterium
728	6	1.5	62	21	AAU52600	Arabidopsis thalia	801	6	1.5	78	22	AA025188	Peptide #4457 enco
729	6	1.5	63	22	AA008532	Human polypeptide	802	6	1.5	78	22	AA025188	Peptide #4543 enco
730	6	1.5	63	23	ABP07828	Human ORFX protein	803	6	1.5	78	22	AA025188	Protein #4351 enco
731	6	1.5	64	22	AA008423	Human ORFX protein	804	6	1.5	78	22	AA025188	Human brain expres
732	6	1.5	64	23	ABP34547	Simian (sal1) rota	805	6	1.5	78	22	AA025188	Human bone marrow
733	6	1.5	65	20	AAU13119	Human secreted pro	806	6	1.5	78	22	AA025188	Peptide #4436 enco
734	6	1.5	65	22	AA067537	Human acid sequenc	807	6	1.5	78	22	AA025188	Peptide #4550 enco
735	6	1.5	65	23	ABP35016	Human ORF3989 prot	808	6	1.5	78	22	AA025188	Peptide #4325 enco
736	6	1.5	66	18	AAW27858	Staphylococcus aur	809	6	1.5	78	22	AA025188	Human peptide enco
737	6	1.5	66	22	AA090086	C glutamicum prote	810	6	1.5	79	20	AA025188	Chlamydia pneumoni
738	6	1.5	66	22	AA067539	Amino acid sequenc	811	6	1.5	79	20	AA025188	Human liver peptid
739	6	1.5	66	23	ABP39721	Staphylococcus epi	812	6	1.5	79	22	ABG55647	

813	6	1.5	79	22	AAU62945	Proionibacterium
814	6	1.5	79	22	ABG24008	Novel human diagno
815	6	1.5	79	22	ABBI15391	Human nervous syst
816	6	1.5	79	22	AAU73896	Human bone marrow
817	6	1.5	79	22	AAU34075	Peptide #8112 enco
818	6	1.5	79	22	ABG43784	Human peptide enco
819	6	1.5	79	23	ABP33869	Human ORF2842 prot
820	6	1.5	80	20	ABP05241	Human ORFX protein
821	6	1.5	80	20	AAV01439	Secreted protein e
822	6	1.5	80	21	AAV87242	Human signal pepti
823	6	1.5	80	23	ABP62911	Human polypeptide
824	6	1.5	81	20	AAV60236	Human endometrium
825	6	1.5	81	22	ABBS59810	Drosophila melanog
826	6	1.5	81	22	ABB67211	Drosophila melanog
827	6	1.5	81	22	AAU28107	Novel human secret
828	6	1.5	81	23	AAO2800	Protein of drug me
829	6	1.5	81	23	ABP07049	Human ORFX protein
830	6	1.5	82	21	AAU26629	Arabidopsis thalia
831	6	1.5	82	22	AAU66762	Proionibacterium
832	6	1.5	82	22	AAU32178	Novel human secret
833	6	1.5	82	22	AAO06615	Human polypeptide
834	6	1.5	83	21	AAU24280	Arabidopsis thalia
835	6	1.5	83	21	AAO3703	Human secreted pro
836	6	1.5	83	22	ABG05602	Novel human diagno
837	6	1.5	83	22	AAO12440	Human polypeptide
838	6	1.5	83	22	AAO12487	Human polypeptide
839	6	1.5	83	23	ABP31783	Human ORF756 prote
840	6	1.5	84	20	AAV04950	Human ORFX protein
841	6	1.5	84	21	AAV13274	Mycobacterium spec
842	6	1.5	84	21	AAV1328	Human secreted pro
843	6	1.5	84	21	AAV1328	Zea mays protein f
844	6	1.5	84	21	AAU25673	Arabidopsis thalia
845	6	1.5	84	21	AAU54842	Arabidopsis thalia
846	6	1.5	84	22	AAO11259	Human polypeptide
847	6	1.5	85	21	AAU51997	Arabidopsis thalia
848	6	1.5	85	22	ABG20494	Novel human diagno
849	6	1.5	85	22	AAU84189	Human immune/haema
850	6	1.5	85	22	AAU84189	Human immune/haema
851	6	1.5	85	22	AAU84189	B. subtilis surfac
852	6	1.5	85	22	AAU84189	B. subtilis surfac
853	6	1.5	85	23	ABP40555	Staphylococcus epi
854	6	1.5	85	23	ABP02976	Human ORFX protein
855	6	1.5	85	24	ABP71850	Human endothelial
856	6	1.5	86	21	AAU32570	Eucalyptus grandis
857	6	1.5	86	22	ABG01722	Novel human diagno
858	6	1.5	86	22	ABG29632	Novel human diagno
859	6	1.5	86	22	AAU82819	Human immune/haema
860	6	1.5	86	23	ABP79205	Human prostate spe
861	6	1.5	86	23	ABP05717	Human ORFX protein
862	6	1.5	87	22	AAO19970	Human G-protein su
863	6	1.5	87	22	AAO08505	Human polypeptide
864	6	1.5	87	22	AAU89759	C glutathione prote
865	6	1.5	87	22	AAU82188	MMV p24 protein c
866	6	1.5	87	22	AAU82188	Mason-Pfizer monke
867	6	1.5	87	22	AAU82188	Human secreted pro
868	6	1.5	87	22	AAU82188	Human secreted pro
869	6	1.5	87	23	ABG60107	Human DIRP polype
870	6	1.5	88	22	AAU94877	Proionibacterium
871	6	1.5	88	22	AAU94352	Human reproductive
872	6	1.5	88	22	AAU94352	Human breast or ov
873	6	1.5	89	22	AAU59633	Proionibacterium
874	6	1.5	89	22	AAU59633	Novel human secret
875	6	1.5	89	23	ABP09354	Human ORFX protein
876	6	1.5	90	21	AAU13884	Arabidopsis thalia
877	6	1.5	90	21	AAU51275	Arabidopsis thalia
878	6	1.5	90	21	AAU51275	Human secreted pro
879	6	1.5	90	21	AAU51275	Human secreted pro
880	6	1.5	90	22	AAU51275	Human polymetric im
881	6	1.5	90	22	AAU51275	Novel human colon
882	6	1.5	90	22	AAU51275	Human digestive sy
883	6	1.5	90	22	AAU51275	Human polypeptide
884	6	1.5	90	23	AAU51275	Human lung-specifi
885	6	1.5	91	22	AAU51275	Human PDZPI domain
886	6	1.5	91	22	AAU51275	Proionibacterium
887	6	1.5	91	22	AAU51275	Lipid biosynthesis
888	6	1.5	91	22	AAU51275	P. patens lipid met
889	6	1.5	91	23	AAU51275	Human PDZ domain #
890	6	1.5	93	22	AAU51275	Novel human secret
891	6	1.5	93	22	AAU51275	Human immune/haema
892	6	1.5	94	20	AAU51275	Helicobacter pylori
893	6	1.5	94	20	AAU51275	Helicobacter pylori
894	6	1.5	94	21	AAU51275	Human prostate can
895	6	1.5	94	21	AAU51275	Eucalyptus grandis
896	6	1.5	94	21	AAU51275	Arabidopsis thalia
897	6	1.5	94	21	AAU51275	Arabidopsis thalia
898	6	1.5	94	22	AAU51275	Novel human diagno
899	6	1.5	94	22	AAU51275	Human secreted pro
900	6	1.5	94	22	AAU51275	Novel human colon
901	6	1.5	94	22	AAU51275	Novel human secret
902	6	1.5	94	22	AAU51275	Novel human secret
903	6	1.5	94	22	AAU51275	Human immune/haema
904	6	1.5	94	24	AAU51275	Human digestive sy
905	6	1.5	95	21	AAU51275	Human polypeptide
906	6	1.5	95	21	AAU51275	Human cancer assoc
907	6	1.5	95	21	AAU51275	Arabidopsis thalia
908	6	1.5	95	21	AAU51275	Arabidopsis thalia
909	6	1.5	95	21	AAU51275	Arabidopsis thalia
910	6	1.5	95	22	AAU51275	Novel human secret
911	6	1.5	95	23	AAU51275	Novel human secret
912	6	1.5	96	18	AAU51275	Leishmania major S
913	6	1.5	96	18	AAU51275	Leishmania major S
914	6	1.5	96	22	AAU51275	Proionibacterium
915	6	1.5	96	22	AAU51275	Novel human secret
916	6	1.5	96	22	AAU51275	Human reproductive
917	6	1.5	96	22	AAU51275	Human breast or ov
918	6	1.5	96	22	AAU51275	Human ORFX protein
919	6	1.5	97	22	AAU51275	Novel human diagno
920	6	1.5	97	22	AAU51275	Novel human secret
921	6	1.5	97	22	AAU51275	Human immune/haema
922	6	1.5	97	23	AAU51275	Human topoisomeras
923	6	1.5	98	23	AAU51275	Human polypeptide
924	6	1.5	98	23	AAU51275	Human signal pepti
925	6	1.5	99	22	AAU51275	Amino acid sequenc
926	6	1.5	100	22	AAU51275	Novel human secret
927	6	1.5	100	22	AAU51275	N. gonorrhoeae ami
928	6	1.5	101	24	AAU51275	Human liver peptid
929	6	1.5	101	22	AAU51275	Drosophila melanog
930	6	1.5	101	22	AAU51275	Peptide #3965 enco
931	6	1.5	101	22	AAU51275	Protein #3801 enco
932	6	1.5	101	22	AAU51275	Human natriuretic
933	6	1.5	101	22	AAU51275	Human bone marrow
934	6	1.5	101	22	AAU51275	Peptide #3994 enco
935	6	1.5	101	22	AAU51275	P. patens S-adenos
936	6	1.5	101	23	AAU51275	Human secretory po
937	6	1.5	101	23	AAU51275	Human secretory po
938	6	1.5	101	23	AAU51275	Human secretory po
939	6	1.5	101	23	AAU51275	Human secretory po
940	6	1.5	101	24	AAU51275	Human secretory po
941	6	1.5	101	24	AAU51275	Human secretory po
942	6	1.5	102	20	AAU51275	Human secretory po
943	6	1.5	102	20	AAU51275	Human secretory po
944	6	1.5	102	22	AAU51275	Novel human diagno
945	6	1.5	102	22	AAU51275	Human polypeptide
946	6	1.5	102	22	AAU51275	Human protein sequ
947	6	1.5	103	22	AAU51275	Novel human diagno
948	6	1.5	103	22	AAU51275	Novel human secret
949	6	1.5	103	22	AAU51275	Human immune/haema
950	6	1.5	103	22	AAU51275	Human EST encoded
951	6	1.5	103	22	AAU51275	Human polypeptide
952	6	1.5	103	23	AAU51275	Human K-beta2 pro
953	6	1.5	103	23	AAU51275	E. coli strain K12
954	6	1.5	104	21	AAU51275	Human ORFX ORF680
955	6	1.5	104	21	AAU51275	Arabidopsis thalia
956	6	1.5	104	22	AAU51275	Novel human diagno
957	6	1.5	104	23	AAU51275	Human secretory pr
958	6	1.5	104	23	AAU51275	Human ORFX protein

959	6	1.5	105	18	AAW25792	Leishmania major S
960	6	1.5	105	22	AAU64044	Procionbacterium
961	6	1.5	106	22	AAU44465	Procionbacterium
962	6	1.5	107	21	AAU10493	N. meningitidis BA
963	6	1.5	107	21	AAU74672	Neisseria meningit
964	6	1.5	107	21	AAU74673	Neisseria meningit
965	6	1.5	107	22	AAU63357	Human breast cancer
966	6	1.5	107	22	AAU63357	Human breast cancer
967	6	1.5	107	22	AAU63357	Human breast cancer
968	6	1.5	107	24	AAU99939	Breast specific po
969	6	1.5	108	21	AAU14592	Arabidopsis thalia
970	6	1.5	108	21	AAU00317	Human secreted pro
971	6	1.5	108	22	AAU15477	Novel human diagno
972	6	1.5	108	22	AAU85779	Human immune/haema
973	6	1.5	109	15	AAU66641	Human breast speci
974	6	1.5	109	15	AAU51504	Plin protein Vari
975	6	1.5	109	22	AAU92415	C glutaminc prote
976	6	1.5	109	22	AAU76631	Corynebacterium gl
977	6	1.5	109	23	AAU10815	Human ORF protein
978	6	1.5	110	22	AAU03790	Human polypeptide
979	6	1.5	110	23	AAU82188	S. typhimurium cyt
980	6	1.5	111	22	AAU14220	Novel human diagno
981	6	1.5	111	24	AAU80244	N. gonorrhoeae ami
982	6	1.5	112	21	AAU12364	Fragment of human
983	6	1.5	112	22	AAU32786	Novel human secret
984	6	1.5	113	22	AAU32786	Novel human secret
985	6	1.5	114	22	AAU00257	Human polypeptide
986	6	1.5	114	24	AAU79769	N. gonorrhoeae ami
987	6	1.5	115	18	AAU20848	H. pylori cytoplas
988	6	1.5	115	23	AAU70160	Human prey protein
989	6	1.5	116	21	AAU42417	Human ORF ORF2181
990	6	1.5	116	22	AAU04138	Human polypeptide
991	6	1.5	116	23	AAU99941	Human transcrip
992	6	1.5	116	23	AAU08252	Human ORF protein
993	6	1.5	117	22	AAU99646	Human excretory re
994	6	1.5	117	22	AAU42461	Human kidney relat
995	6	1.5	118	20	AAU16956	Human fascicular a
996	6	1.5	118	20	AAU12314	Human 5' EST seque
997	6	1.5	118	22	AAU61583	Propionibacterium
998	6	1.5	118	23	AAU57818	Human synaptic ves
999	6	1.5	119	22	AAU87677	Novel central nerv
1000	6	1.5	119	22	AAU68016	Propionibacterium
					AAU04440	Human polypeptide

ALIGNMENTS

RESULT 1
AAU90283
ID AAU90283 standard; Protein; 412 AA.
XX
AC AAU90283;
XX
DT 24-OCT-2000 (first entry)
XX
DE N. meningitidis BASB05 protein sequence.
XX
KW BASB05; diagnosis; microbial infection; invasive bacterial disease;
KW Neisseria meningitidis infection; upper respiratory tract infection;
KW bacteraemia; meningitis; therapy.
XX
OS Neisseria meningitidis.
XX
PN WO200043517-A1.
XX
PD 27-JUL-2000.
XX
PF 19-JAN-2000; 2000WO-EP00425.
XX
PR 22-JAN-1999; 99GB-0001462.
XX
PR 29-JAN-1999; 99GB-0002069.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonard J;
PI WPI: 2000-476199/41.
XX N-PSDB; AAA37645.
DR
XX Isolated BASB05 polypeptides, polynucleotides, and antibodies, the
PT polypeptides and polynucleotides are useful as vaccines for treating
PT and diagnosing a microbial infection such as a Neisseria meningitidis
PT infection
XX
PS Claim 3, Page 63; 82pp; English.
XX
XX This sequence represents the Neisseria meningitidis BASB05 polypeptide
CC of the invention. The BASB05 polypeptides and polynucleotides are useful
CC for diagnosing and treating microbial infections such as a Neisseria
CC meningitidis infection. They can also be used to treat any disease caused
CC by or related to infection by a bacteria, including upper respiratory
CC tract infection, invasive bacterial diseases (such as bacteraemia) and
CC meningitis.
XX
SQ Sequence 412 AA;
XX
XX Query Match 100.0%; Score 412; DB 21; Length 412;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MAFVAFKMRRAAALAAVALVLSGCKGDAAGSGPAGREAPAVGVTVHPQVALT 60
DB 1 MAFVAFKMRRAAALAAVALVLSGCKGDAAGSGPAGREAPAVGVTVHPQVALT 60
QY 61 VELPRLSLRTADVRAVGGIIQKRLFOESYVRAQPLYOIDSITYEAMLESARAQLA 120
DB 61 VELPRLSLRTADVRAVGGIIQKRLFOESYVRAQPLYOIDSITYEAMLESARAQLA 120
QY 121 TQAATLAAADADLARYKPLVAEAVSROEYPAATYAKRSAGVKAQAATKSAGINLR 180
DB 121 TQAATLAAADADLARYKPLVAEAVSROEYPAATYAKRSAGVKAQAATKSAGINLR 180
QY 181 SRITPPIGFGTQGSVSEGTLLNMGDTVLTATITQTNMYNVNQSASEWKLRRQIAEG 240
DB 181 SRITPPIGFGTQGSVSEGTLLNMGDTVLTATITQTNMYNVNQSASEWKLRRQIAEG 240
QY 241 KLLADGVIAVGIRKDDGVYPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYV 300
DB 241 KLLADGVIAVGIRKDDGVYPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYV 300
QY 301 RVLMDQVAVDNAFVVPQAVTRGAKDTVMIVNAQCGMEPREVTVAQOOGTMMIVTSGIKD 360
DB 301 RVLMDQVAVDNAFVVPQAVTRGAKDTVMIVNAQCGMEPREVTVAQOOGTMMIVTSGIKD 360
QY 361 GDKVVEGISIAGITGAKKVPKEWASSENOAAAPQSGVQTASBAKTASEA 412
DB 361 GDKVVEGISIAGITGAKKVPKEWASSENOAAAPQSGVQTASBAKTASEA 412
RESULT 2
AAU91064
ID AAU91064 standard; Protein; 412 AA.
XX
AC AAU91064;
XX
DT 05-JUN-2002 (first entry)
XX
DE Neisseria cell surface polypeptide #2.
XX
KW Cell surface protein; antibacterial; antimicrobial.
XX
OS Neisseria meningitidis.
XX
PN WO200216612-A2.
XX
PD 28-FEB-2002.

```

XX 21-AUG-2001; 2001WO-GB03759.
XX
XX 24-AUG-2000; 2000GB-0020952.
XX
XX (MIGR-) MICROSCIENCE LTD.
XX
XX Lane JD, Hughes MJG, Santangelo JD;
XX
XX WPI: 2002-280941/32.
XX N-PSDB; ABK54080.
XX
XX Novel peptide encoded by Neisseria meningitidis, useful for manufacture
XX of medicament for treatment or prevention of condition associated with
XX infection by Neisseria or Gram-negative bacteria -
XX
XX Claim 5, Page 19-20; 79pp; English.
XX
XX The invention relates to polypeptides located on the cell surface of
XX Neisseria meningitidis, and the polynucleotides encoding them. The
XX sequences of the invention are useful for therapeutic or diagnostic use,
XX in the manufacture of a medicament for use in treatment or prevention of
XX a condition associated with infection by Neisseria or Gram-negative
XX bacteria. The sequences are also useful for screening potential
XX antimicrobial drugs or for detection of virulence. Sequences
XX AAU91063-AAU91079 represent Neisseria meningitidis polypeptides of the
XX invention.
XX
XX Sequence 412 AA;
SQ
XX
XX Query Match 32.3%; Score 133; DB 23; Length 412;
XX Best Local Similarity 100.0%; Pred. No. 2.6e-129;
XX Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 274 VNESIGQITLRAAVPNDQNIIMPGIYVRVLMQVAVDNAFVFPQQAIVTGADQTMIVNA 333
XX |||||
XX 274 VNESIGQITLRAAVPNDQNIIMPGIYVRVLMQVAVDNAFVFPQQAIVTGADQTMIVNA 333
XX
XX 334 OCGMEPREVTVAQOOGTMMIVTSGKDGPKVVEGISTAGTGAKKVTPKEMASSENOAA 393
XX |||||
XX 334 OCGMEPREVTVAQOOGTMMIVTSGKDGPKVVEGISTAGTGAKKVTPKEMASSENOAA 393
XX
XX 394 APQSGVQTASEAK 406
XX |||||
XX 394 APQSGVQTASEAK 406
XX
XX Db
XX
XX RESULT 3
XX ABP78092
XX ID ABP78092 standard; Protein; 412 AA.
XX
XX AC ABP78092;
XX
XX DT 07-MAR-2003 (first entry)
XX
XX DE N. gonorrhoeae amino acid sequence SEQ ID 2714.
XX
XX KW Antibacterial; infection; vaccine; gene therapy.
XX
XX OS Neisseria gonorrhoeae.
XX
XX PN WO200279243-A2.
XX
XX PD 10-OCT-2002.
XX
XX PF 12-FEB-2002; 2002WO-IB02069.
XX
XX PR 12-FEB-2001; 2001GB-0003424.
XX
XX PA (CHIR-) CHIRON SPA.
XX
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX

```

```

DR WPI: 2003-058415/05.
DR N-PSDB; ABZ39062.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection -
XX
XX Disclosure; Page 386; 815pp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention.
XX
XX Sequence 412 AA;
SQ
XX
XX Query Match 24.3%; Score 100; DB 24; Length 412;
XX Best Local Similarity 100.0%; Pred. No. 5.9e-95;
XX Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 148 QEYDAAVTAKSAEKGVAQAQAIKSAGININRSRITAPISGFIQSKVSEGTLLMAGDT 207
XX |||||
XX 148 QEYDAAVTAKSAEKGVAQAQAIKSAGININRSRITAPISGFIQSKVSEGTLLMAGDT 207
XX
XX 208 TVLATIRQTNPVYVNTQASAEVWKLRRQIABGKLLADG 247
XX |||||
XX 208 TVLATIRQTNPVYVNTQASAEVWKLRRQIABGKLLADG 247
XX
XX Db
XX
XX RESULT 4
XX AAY70472
XX ID AAY70472 standard; Protein; 212 AA.
XX
XX AC AAY70472;
XX
XX DT 21-JUN-2000 (first entry)
XX
XX DE Human p53 target molecule, PRG6 protein.
XX
XX KW PRG6; p53 target; human; modulate; cell proliferation; immunomodulatory;
XX chromosome 19p13.2-13.1; cytostatic; gene therapy; tumour cell; inducer;
XX diagnosis; therapeutic; proliferative disease; cell cycle arrest; cancer;
XX treatment; apoptosis; knockout animal; cancer susceptibility; FOU-domain;
XX homeo-domain.
XX
XX KW Homo sapiens.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX
XX FT Misc-difference 66 /note= "Encoded by TTG"
XX
XX FT Misc-difference 202 /note= "Encoded by CCA"
XX
XX PN WO200012526-A1.
XX
XX PD 09-MAR-2000.
XX
XX PF 27-AUG-1999; 99WO-US19551.
XX
XX PR 28-AUG-1998; 98US-0098251.
XX
XX PA (UYPR-) UNIV PRINCETON.
XX
XX PI Horikoshi N, Shenk T;
XX
XX DR WPI: 2000-246724/21.
XX N-PSDB; AAZ51674.
XX
XX New p53-inducible isolated nucleic acid molecule including open reading
XX frame encoding human homolog of Drosophila melanogaster peroxidasin,
XX

```

PT useful e.g. in detection and treatment of cancer -
XX
XX
PS Disclosure; Fig 9; 83pp; English.
XX
CC The present sequence is the protein encoded by PRG6 gene, that is
CC decreased in response to induction of p53 activity in human colon
CC cancer EBI cells. Repression of PRG6 is by a liver specific factor. It
CC is a basic protein localised in the nucleus that has POU-domain like
CC sequence and two homeo-domain like sequences. It may function as a
CC transcription factor. They are potential targets of p53 regulatory
CC activity and are useful for modulation of cellular proliferation. PRG6
CC gene is localised to human chromosome 19p13.2-13.1. The PRG target
CC molecules have cytostatic and immunomodulatory activity. PRG nucleotides,
CC proteins and antibodies are useful as diagnostic and therapeutic agents
CC for detection and treatment of cancer and other proliferative diseases.
CC The gene/cDNA may be used for gene therapy, to restore a gene function
CC downstream of p53, that cannot be activated in the p53-deficient tumour
CC cell. Antibodies can be used as inducers of cell cycle arrest and/or
CC apoptosis. The DNA sequences can be used to generate 'knockout' animals
CC as a model of cancer susceptibility.
XX
SQ Sequence 212 AA;
Query Match 2.4%; Score 10; DB 21; Length 212;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 RAALAAAVA 19
DB 192 RAALAAAVA 201
RESULT 5
AAG93326
ID AAG93326 standard; Protein; 222 AA.
XX
AC AAG93326;
XX
DT 13-SEP-2001 (first entry)
XX
DE Human protein HP10438.
XX
KM Human; gene therapy; tumour.
XX
OS Homo sapiens.
XX
PN WO200142302-A1.
XX
PD 14-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-JP06531.
XX
PR 06-DEC-1999; 99JP-0346663.
XX
PR 06-DEC-1999; 99JP-0346664.
XX
PR 08-FEB-2000; 2000JP-0031062.
XX
PR 10-FEB-2000; 2000JP-0034090.
XX
PR 10-FEB-2000; 2000JP-0034091.
XX
PR 14-FEB-2000; 2000JP-0035829.
XX
PR 14-FEB-2000; 2000JP-0035899.
XX
PR 14-MAR-2000; 2000JP-0071161.
XX
PR 30-MAY-2000; 2000JP-0160851.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Kato S, Eguuchi C, Saeki M;
XX
DR MPI; 2001-381646/40.
XX
DR N-PSDB; AAH68611.
XX
PT Human protein originated from tumor cell line, applicable as drug,
PT reagent for studying intracellular protein networks and protein source
PT for drug screening, also encoded cDNA for gene diagnosis and gene
PT therapy -

XX
PS Claim 1; Pages 451-452; 471pp; Japanese.
XX
XX
CC The present sequence is a human protein. The human protein, preferably
CC originated from tumor cell line, is applicable as a drug, a reagent for
CC studying intracellular protein networks and a protein source for
CC screening proteins for binding low molecular weight drugs. The human
CC protein coding sequence is useful for gene diagnosis and gene therapy,
CC expression vectors and transformant cells for detection of ligands and
CC receptors.
XX
SQ Sequence 222 AA;
Query Match 2.4%; Score 10; DB 22; Length 222;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 RAALAAAVA 19
DB 202 RAALAAAVA 211
RESULT 6
AAU61543
ID AAU61543 standard; Protein; 448 AA.
XX
AC AAU61543;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #22439.
XX
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200161581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
XX
PR 02-JUN-2000; 2000US-208841P.
XX
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR MPI; 2001-616774/71.
XX
DR N-PSDB; AAS59618.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 22738; 1069pp; English.
XX
XX
CC Sequences AAU939105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 448 AA;

Query Match 2.2%; Score 9; DB 22; Length 448;

Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 LKDGKVVV 366
 147 LKDGKVVV 155

RESULT 7

ID ABP76681 standard; Protein; 19938 AA.

XX ABP76681;

DT 26-FEB-2003 (first entry)

DE Streptomyces viridochromogenes Av1 gene cluster polypeptide frame 5.

KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;

KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.

OS Streptomyces viridochromogenes.

PN WO200268436-A1.

PD 06-SEP-2002.

PF 24-AUG-2001; 2001WO-EP09815.

PR 25-FEB-2001; 2001DE-1009166.

PA (COMB-) COMBINATURE BIOPHARM AG.

PI Weinbauer G, Muehlenweg A, Trefzer A, Bechtold A;

DR WPI; 2003-018650/01.

DR N-PSDB; AB237516.

PT New avilamycin derivatives, useful for treatment of infections, and

PT nucleic acid encoding avilamycin synthesis enzymes -

PS Example 1; Page 68-301; 319pp; German.

XX The invention relates to avilamycin derivatives (I) with antibacterial,
 CC virucide, protozoacide and fungicide activity. (I) are useful for
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human
 CC or veterinary medicine, particularly where caused by Staphylococcus
 CC aureus. (I) are more hydrophilic than known avilamycins. The present
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
 CC viridochromogenes Avilamycin A biosynthetic gene cluster
 CC (AB237515-AB237516).

XX Sequence 19938 AA;

Query Match 2.2%; Score 9; DB 24; Length 19938;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAAVA 19

Db 13793 AAALAAVA 13801

RESULT 8

ID AAY37704 standard; Protein; 111 AA.

XX AAY37704;

DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis ribosomal polypeptide.

KW Vaccine; eye disease; conventional trachoma; nongonococcal trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;

KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW bartolinitis; pneumonia; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

PN WO928475-A2.

PD 10-JUN-1999.

PF 27-NOV-1998; 98WO-IB01939.

PR 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

PA (GEST) GENSET.

PI Griffiths R;

DR WPI; 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis

PS Disclosure; Page 1317; 1755pp; English.

CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nongonococcal trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perihhepatitis, bartolinitis; pneumonia; venereal lymphogranulomatosis;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX Sequence 111 AA;

Query Match 1.9%; Score 8; DB 20; Length 111;

Best Local Similarity 100.0%; Pred. No. 9.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 DKVVEGI 369
 32 DKVVEGI 39

RESULT 9

ID AAG98889 standard; Protein; 171 AA.

XX AAG98889;

DT 26-SEP-2001 (first entry)

DE E. coli growth and proliferation related protein sequence SEQ ID NO:359.

```

XX Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism.
XX
XX Escherichia coli.
XX
XX WO200134810-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30950.
XX
XX 09-NOV-1999; 99US-0164415.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Forsyth RA, Ohlsen K, Zyskind J;
XX
XX WPI; 2001-335933/35.
XX
XX N-PSDB; AAH84560.
XX
XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful
XX for screening for homologous genes and for designing expression vectors
XX
XX Claim 19; Page 435; 522pp; English.
XX
XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
XX related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
XX growth and proliferation related proteins given in AAG99078 and AAG98830
XX to AAG99899. (1) can be used as potential targets for the generation of
XX new antimicrobial agents, and for identification of compounds which
XX interact with the gene products of (1). In addition the expression of
XX (1) and the purification of the proteins, the purified proteins can be
XX used to generate reagents and screen small molecule libraries or other
XX candidate compound libraries for compounds that can be further developed
XX to yield novel antimicrobial compounds. In addition, nucleic acid probes
XX complementary to (1) that are specific for particular species of
XX microorganisms can be used to identify particular microorganism species
XX in clinical specimens, therefore, providing a rapid and dependable
XX method by which to identify the causative agents of a bacterial
XX infection. Also, antibodies generated against proteins translated from
XX cDNA transcribed from proliferation-required sequences can also be used
XX to screen for specific microorganisms that produce such proteins in a
XX species-specific manner. AAH84371 and AAH84670 represent sequencing
XX primers used in the isolation of E. coli growth and proliferation
XX related sequence, which are used in an example from the present
XX invention.
XX
XX Sequence 171 AA;
XX
XX Query Match 1.9%; Score 8; DB 22; Length 171;
XX Best Local Similarity 100.0%; Pred. No. 15;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 201 LUNAGDTT 208
XX |||||
XX Db 61 LUNAGDTT 68
XX
XX RESULT 10
XX ABB05926
XX ID ABB05926 standard; Protein; 181 AA.
XX
XX ABB05926;
XX
XX 08-APR-2003 (first entry)
XX
XX M. tuberculosis and M. leprae marker protein #577.
XX
XX Mycobacteriosis; survival; virulence; protective antigen; vaccine;
XX mycobacterial disease; tuberculosis; leprosy.
XX

```

```

OS Mycobacterium tuberculosis.
OS Mycobacterium leprae.
XX
XX WO200274903-A2.
XX
XX 26-SEP-2002.
XX
XX 22-FEB-2002; 2002WO-IB01973.
XX
XX 22-FEB-2001; 2001US-270123P.
XX
XX (INSP) INST PASTEUR.
XX
XX Cole S;
XX
XX WPI; 2002-759865/82.
XX
XX Identifying and selecting genes for survival or virulence of
XX mycobacteria by a comparative genomic analysis of the sequences of
XX Mycobacterium tuberculosis and M. leprae -
XX
XX Claim 17; Page 791; 874pp; English.
XX
XX This invention relates to a novel method for identifying essential genes
XX for survival or virulence of mycobacteria species. The method comprises
XX aligning the genomic sequence of a first mycobacterium species on a
XX genomic sequence of a second mycobacterium species and selecting a
XX polynucleotide sequence that is highly conserved in both genomes with no
XX counterparts in other bacterial genomic sequences and that corresponds
XX to an essential gene for the survival or virulence of mycobacterium
XX species. The method of the invention is useful for detecting M.
XX tuberculosis or M. leprae infection. The method reduces the number of
XX potential new targets and protective antigens for new drugs and vaccine
XX compositions to treat and prevent mycobacterial diseases, particularly
XX tuberculosis and leprosy. The present sequence represents a marker
XX protein from Mycobacterium tuberculosis and Mycobacterium leprae
XX identified using the method of the invention.
XX
XX Sequence 181 AA;
XX
XX Query Match 1.9%; Score 8; DB 23; Length 181;
XX Best Local Similarity 100.0%; Pred. No. 16;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 11 AAALAAV 18
XX |||||
XX Db 16 AAALAAV 23
XX
XX RESULT 11
XX AAB34400
XX ID AAB34400 standard; Protein; 185 AA.
XX
XX AAB34400;
XX
XX 26-JAN-2001 (first entry)
XX
XX Gene 30 human secreted protein homologous amino acid sequence #161.
XX
XX Human; secreted protein; diagnosis; neuroprotective; cyostatic;
XX cardioactive; immunomodulatory; muscular active general; vulnerrary;
XX gastrointestinal; nephrotoxic; antiinfective; gynaecological; and
XX antibacterial; gene therapy's detection; cancer; chromosome marker;
XX chromosome identification; neural disorder; immune disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; wound healing; infectious disease; preservative;
XX food additive.
XX
XX Caenorhabditis elegans.
XX
XX OS
XX OS
XX PN WO200056883-A1.
XX

```



```

PD 28-SEP-2000.
XX
XX
PF 16-MAR-2000; 2000WO-US06822.
XX
XX 23-MAR-1999; 99US-0126054.
PR 10-DEC-1999; 99US-0169916.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-587666/55.
XX
XX Human secreted proteins and gene sequences encoding them, useful for
XX detecting, preventing, and treating disorders such as cancer,
XX neurological disorders and immune system disorders -
XX
XX Disclosure; Page 409-410, 429pp; English.
XX
XX The polynucleotide sequences given in AAC59566 to AAC59614 encode the
XX human secreted proteins given in AAB34299 to AAB34347. AAB34348 to
XX AAB34437 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Example of activities include:
XX neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular
XX active general; vulnery; gastrointestinal; nephrotropic;
XX antifetive; gynaecological; and antibacterial. The polynucleotides
XX can be used for the detection of various disorders such as cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The secreted proteins can be used to
XX treat disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wound healing, and infectious diseases. The proteins can
XX also be used as a food additive or preservative to increase or decrease
XX storage capabilities. AAC59557 to AAC59565 and AAB34238 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 185 AA;
XX
XX Query Match 1.9%; Score 8; DB 21; Length 185;
XX Best Local Similarity 100.0%; Pred. No. 16;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 239 EGKLAAD 246
XX |||||
XX 141 EGKLAAD 148
XX
XX RESULT 12
XX AAY86471
XX ID AAY86471 standard; Protein; 194 AA.
XX
XX AAY86471;
XX
XX 19-APR-2000 (first entry)
XX
XX Human gene 51-encoded protein fragment, SEQ ID NO:386.
XX
XX Human; secreted protein; cancer; tumour; developmental abnormality;
XX foetal deficiency; blood disorder; immune system disorder; inflammation;
XX autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
XX schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
XX atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
XX digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
XX therapy; chromosome 19.
XX
XX Homo sapiens.
XX
XX WO9966041-A1.
XX
XX 23-DEC-1999.
XX
XX

```

```

PF 15-JUN-1999; 99WO-US13418.
XX
XX 16-JUN-1998; 98US-0089507.
XX 16-JUN-1998; 98US-0089508.
PR 16-JUN-1998; 98US-0089509.
PR 16-JUN-1998; 98US-0089510.
PR 22-JUN-1998; 98US-0090112.
XX 22-JUN-1998; 98US-0090113.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA,
XX Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;
XX Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;
XX
XX WPI; 2000-106100/09.
XX
XX New isolated human genes and the secreted polypeptides they encode,
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders -
XX
XX Disclosure; Page 103, 586pp; English.
XX
XX AA297019 to AA297137 represent 94 isolated human secreted protein genes.
XX AAY86215 to AAY86333 are the secreted proteins encoded by the 94 human
XX genes. This sequence represents a fragment of one of the human secreted
XX proteins. The genes and their corresponding secreted polypeptides are
XX useful for preventing, treating or ameliorating medical conditions,
XX e.g., by protein or gene therapy. Also pathological conditions can be
XX diagnosed by determining the amount of the new polypeptides in a sample
XX or by determining the presence of mutations in the new genes. Specific
XX uses are described for each of the 94 genes, based on which tissues they
XX are most highly expressed in, and include developing products for the
XX diagnosis or treatment of cancer, tumours, developmental abnormalities
XX and foetal deficiencies, blood disorders, diseases of the immune system,
XX autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
XX disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
XX disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
XX disorders, digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The sequences shown in AAY86334 to AAY86585 represent fragments of the
XX secreted proteins.
XX
XX Sequence 194 AA;
XX
XX Query Match 1.9%; Score 8; DB 21; Length 194;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 AALAAVA 19
XX |||||
XX 176 AALAAVA 183
XX
XX RESULT 13
XX ABB54244
XX ID ABB54244 standard; Protein; 306 AA.
XX
XX ABB54244;
XX
XX 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein yJfE.
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis IL1403.
XX
XX FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX

```

XX 11-APR-2000; 2000FR-0004630.
 PR (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PA Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 PI WPI; 2002-043418/06.
 DR New nucleotide sequence useful in the identification or Lactococcus
 PT lactis and related species -
 PS Claim 6; SEQ ID No 946; 2504bp; French.
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (AB90521) and related proteins (AB53300-AB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 SQ Sequence 306 AA;

Query Match 1.9%; Score 8; DB 23; Length 306;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 TILNAGDT 207
 Db 202 TILNAGDT 209

RESULT 14
 ABP65884
 ID ABP65884 standard; Protein; 318 AA.
 AC ABP65884;
 XX
 DT 19-NOV-2002 (first entry)
 DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:628.
 XX
 KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KW anticlathric; antibacterial; inhibitor of Salmonella; detection;
 KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 KW rotavirus; food composition; pharmaceutical composition.
 XX
 OS Bifidobacterium longum.
 XX
 PN EP1227152-A1.
 PD 31-JUL-2002.
 PF 30-JAN-2001; 2001EP-0102050.
 PR 30-JAN-2001; 2001EP-0102050.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 DR WPI; 2002-668397/72.
 XX
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
 PT as a probe or primer for detecting and/or identifying Bifidobacterium
 PT longum in a biological sample -
 PS Claim 3; SEQ ID 628; 80bp; English.
 CC The present invention describes a polynucleotide (I) comprising a

CC sequence of a Bifidobacterium genome selected from the nucleotide
 CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
 CC least 90% identity or which hybridises with the sequences given in
 CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
 CC a fusion protein, comprising a sequence selected from 1097 sequences
 CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
 CC encoding a heterologous polypeptide. (I) has anticlathric and
 CC antibacterial activities, and can be used as an inhibitor of Salmonella.
 CC (I) (which is a probe) is useful for the detection and/or identification
 CC of Bifidobacterium longum in a biological sample. A carrier containing
 CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
 CC can be used for preventing and/or treating diarrhoea brought about by
 CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
 CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
 CC fermented products, ice-creams, fermented cereal based products, milk
 CC based powders, infant formula, pet food or a pharmaceutical composition
 CC selected from tablets, liquid bacterial suspensions, dried oral
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
 CC (I) is useful in DNA arrays or chips to carry out analysis of the
 CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
 CC Bifidobacterium related nucleotide sequences given in the Sequence
 CC listing from the present invention but not mentioned further within the
 CC specification.
 CC N.B. The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied by the
 CC European Patent Office.
 CC
 SQ Sequence 318 AA;

Query Match 1.9%; Score 8; DB 23; Length 318;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALAAV 18
 Db 210 AAALAAV 217

RESULT 15
 ABG28560
 ID ABG28560 standard; Protein; 426 AA.
 AC ABG28560;
 XX
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #28551.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 DR Dmanac RT, Liu C, Tang YT;
 XX
 PT WPI; 2001-639362/73.
 DR N-PSDB; AAS92747.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 20; SEQ ID No 58919; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating

XX disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits to assess biodiversity

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. ABG0010-ABG0377 represent novel human

XX CC Note: The sequence data for this patent did not appear in the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 426 AA;

Query Match 1.9%; Score 8; DB 22; Length 426;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALALAA 17

Db 234 RAALALAA 241

RESULT 16

ABP65347

ID ABP65347 standard; Protein; 442 AA.

XX AC ABP65347;

XX DT 19-NOV-2002 (first entry)

XX DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:91.

XX KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

XX KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;

XX KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;

XX KW rotavirus; food composition; pharmaceutical composition.

XX OS Bifidobacterium longum.

XX PN EP1227152-A1.

XX PD 31-JUL-2002.

XX PE 30-JAN-2001; 2001EP-0102050.

XX PR 30-JAN-2001; 2001EP-0102050.

XX PA (NEST) SOC PROD NESTLE SA.

XX PI WPI, 2002-668397/72.

XX DR Novel polynucleotide comprising Bifidobacterium genome sequence useful

XX PT as a probe or primer for detecting and/or identifying Bifidobacterium

XX PT longum in a biological sample

XX PS Claim 3; SEQ ID 91; 80pp; English.

XX The present invention describes a polynucleotide (I) comprising a

XX CC sequence of a Bifidobacterium genome selected from the nucleotide

CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

CC least 90% identity or which hybridises with the sequences given in

CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding

CC a fusion protein, comprising a sequence selected from 1097 sequences

CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide

CC encoding a heterologous polypeptide. (II) has antidiarrheic and

CC antibacterial activities, and can be used as an inhibitor of Salmonella.

CC of Bifidobacterium longum in a biological sample. A carrier containing

CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)

CC can be used for preventing and/or treating diarrhoea brought about by

CC pathogenic bacteria and/or rotavirus. The carrier is a food composition

CC selected from milk, yogurt, curd, cheese, fermented milks, milk based

CC fermented products, ice-creams, fermented cereal based products, milk

CC based powders, infant formula, pet food or a pharmaceutical composition

CC selected from tablets, liquid bacterial suspensions, dried oral

CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.

CC (I) is useful in DNA arrays or chips to carry out analysis of the

CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent

CC Bifidobacterium related nucleotide sequences given in the Sequence

CC listing from the present invention but not mentioned further within the

CC specification.

CC N.B. The sequence data for this patent is not represented in the printed

CC CC specification but is based on sequence information supplied by the

CC CC European Patent Office.

XX SQ Sequence 442 AA;

Query Match 1.9%; Score 8; DB 23; Length 442;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 358 LKQGDQKV 365

Db 128 LKQGDQKV 135

RESULT 17

AAW95959

ID AAW95959 standard; Protein; 478 AA.

XX AC AAW95959;

XX DT 22-JUN-1999 (first entry)

XX DE Human methionine aminopeptidase protein.

XX KW Human; methionine aminopeptidase; expression; protein synthesis system;

XX KW eIF-2-associated glycoprotein; cellular system.

XX OS Homo sapiens.

XX PN US588796-A.

XX PD 30-MAR-1999.

XX PE 18-MAR-1998; 98US-0040705.

XX PR 31-JAN-1996; 96US-0595025.

XX PR 18-MAR-1998; 98US-0040705.

XX PA (U.S.L-) UNIV ST LOUIS.

XX PI Chang Y;

XX DR WPI; 1999-253233/21.

XX DR N-PSDB; AAX29817.

XX PT New polypeptide comprising human methionine aminopeptidase useful

XX PT for monitoring expression in protein synthesis systems

XX PS Claim 1; Fig 1; 11pp; English.

CC This sequence corresponds to a human methionine aminopeptidase.
 CC Over-expression of the protein in a protein synthesis system (in vivo,
 CC in vitro or recombinant) is used to promote protein synthesis by removing
 CC N-terminal methionine from the desired protein. Methionine aminopeptidase
 CC and eIF-2 associated glycoprotein (p67) are shown to be substantially the
 CC same protein and so providing a cellular system with the cloned
 CC nucleotide sequence will serve both cellular functions.

XX
 SQ Sequence 478 AA;

Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 18

ID AAM93215 standard; Protein; 478 AA.

XX AAW93215;

DT 27-MAY-1999 (first entry)

DE Human p67 homologue with methionine aminopeptidase activity protein.

XX Methionine aminopeptidase; eIF-2; protein synthesis; N-terminal; p67;
 KW eukaryotic initiation factor-2 associated glycoprotein; regulatory;
 KM protein modification; N-myristoylation.

XX Homo sapiens.

XX US5885820-A.

XX 23-MAR-1999.

PF 18-MAR-1998; 98US-0040799.

PR 31-JAN-1996; 96US-0595025.

PR 18-MAR-1998; 98US-0040799.

PA (UYSL-) UNITV ST LOUIS.

PI Chang Y;

DR WPI; 1999-228541/19.

DR N-PSDB; AAX22709, AAX22710.

PT Protein with methionine aminopeptidase activity - similar to
 PT eukaryotic initiation factor-2 associated glycoprotein is new
 PS Claim 2; Fig 1; 11pp; English.

XX This sequence represents a human methionine aminopeptidase found to be
 CC similar to eukaryotic initiation factor-2 (eIF-2) associated
 CC glycoprotein (p67). The methionine aminopeptidase protein removes the
 CC N-terminal Met residue from proteins during protein synthesis. This is
 CC essential for subsequent modification of protein, such as in
 CC N-myristoylation. The methionine aminopeptidase is substantially similar
 CC to eukaryotic initiation factor-2 (eIF-2) associated glycoprotein (p67),
 CC and so may facilitate the function of an eukaryotic initiation factor,
 CC thus having a regulatory role in regulation of protein synthesis. The
 CC protein may facilitate protein synthesis by protecting eIF-2 from
 CC phosphorylation. The methionine aminopeptidase polynucleotides can
 CC be used to monitor synthesis of the protein peptidase.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 19

ID AAM94763 standard; protein; 478 AA.

XX AAW94763;

DT 28-APR-1999 (first entry)

DE Mouse type 2 methionine aminopeptidase (MetAP2) putative sequence.

XX Ovalicin; fumagillin; type 2 methionine aminopeptidase; inhibitor;
 KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;
 KM inflammatory disease; immune reaction; autoimmune disease; allergy;
 KW tissue graft rejection; mouse.

XX Mus sp.

OS Key Location/Qualifiers

FT Misc-difference 418

FT label= unknown

PN WO9856372-A1.

PD 17-DEC-1998.

PF 08-JUN-1998; 96WO-US11775.

PR 09-JUN-1997; 97US-0049159.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Griffith EC, Liu JO, Su Z;

DR WPI; 1999-080848/07.

PT New ovalicin and fumagillin derivatives - are inhibitors of type 2
 PT methionine amino-peptidase, useful for treating or diagnosing
 PT diseases involving abnormal angiogenesis or immune reactions

PS Disclosure; Fig 2; 99pp; English.

XX The invention relates to ovalicin and fumagillin derivatives that can
 CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are
 CC useful for treating and/or diagnosing diseases involving abnormal
 CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,
 CC arteriosclerosis) or immune reactions which result in pathology (e.g.
 CC autoimmune disease, allergy and tissue graft rejection). The present
 CC sequence represents a putative amino acid sequence of mouse MetAP2.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
 |||||
 DB 235 NAGDTTVL 242

RESULT 20

ID AAM94764 standard; protein; 478 AA.

XX AAW94764;

XX

DT 28-APR-1999 (first entry)
 XX Rat type 2 methionine aminopeptidase (MetAP2).
 DE
 XX
 XX Ovaliclin; fumagillin; type 2 methionine aminopeptidase; inhibitor;
 KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;
 KW inflammatory disease; immune reaction; autoimmune disease; allergy;
 KW tissue graft rejection; rat.
 OS
 XX Rattus sp.
 XX
 XX WO9856372-A1.
 XX
 XX 17-DEC-1998.
 XX
 XX 08-JUN-1998; 98WO-US11775.
 XX
 XX 09-JUN-1997; 97US-0049159.
 XX
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 XX Griffith EC, Liu JO, Su Z;
 XX
 XX WPI; 1999-080848/07.
 XX
 XX New ovaliclin and fumagillin derivatives - are inhibitors of type 2
 PT methionine amino-peptidase, useful for treating or diagnosing
 PT diseases involving abnormal angiogenesis or immune reactions
 XX
 XX Disclosure; Fig 2; 99pp, English.
 XX
 XX The invention relates to ovaliclin and fumagillin derivatives that can
 CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are
 CC useful for treating and/or diagnosing diseases involving abnormal
 CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,
 CC arteriosclerosis) or immune reactions which result in pathology (e.g.
 CC autoimmune disease, allergy and tissue graft rejection). The present
 CC sequence represents the amino acid sequence of rat MetAP2.
 CC
 SQ Sequence 478 AA;
 Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242
 RESULT 21
 AAM94765
 ID AAM94765 standard; protein; 478 AA.
 XX
 XX AAM94765;
 AC
 XX
 XX 28-APR-1999 (first entry)
 DT
 XX
 XX Human type 2 methionine aminopeptidase (MetAP2).
 DE
 XX Ovaliclin; fumagillin; type 2 methionine aminopeptidase; inhibitor;
 KW MetAP2; angiogenesis; tumour; diabetic retinopathy; arteriosclerosis;
 KW inflammatory disease; immune reaction; autoimmune disease; allergy;
 KW tissue graft rejection; human.
 OS
 XX Homo sapiens.
 XX
 XX WO9856372-A1.
 XX
 XX 17-DEC-1998.
 XX
 XX 08-JUN-1998; 98WO-US11775.
 XX
 XX

PR 09-JUN-1997; 97US-0049159.
 XX
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 XX Griffith EC, Liu JO, Su Z;
 XX
 XX WPI; 1999-080848/07.
 XX
 XX New ovaliclin and fumagillin derivatives - are inhibitors of type 2
 PT methionine amino-peptidase, useful for treating or diagnosing
 PT diseases involving abnormal angiogenesis or immune reactions
 XX
 XX Disclosure; Fig 2; 99pp, English.
 XX
 XX The invention relates to ovaliclin and fumagillin derivatives that can
 CC inhibit type 2 methionine aminopeptidase (MetAP2). These inhibitors are
 CC useful for treating and/or diagnosing diseases involving abnormal
 CC angiogenesis (e.g. tumours, diabetic retinopathy, inflammatory diseases,
 CC arteriosclerosis) or immune reactions which result in pathology (e.g.
 CC autoimmune disease, allergy and tissue graft rejection). The present
 CC sequence represents the amino acid sequence of human MetAP2.
 CC
 SQ Sequence 478 AA;
 Query Match 1.9%; Score 8; DB 20; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242
 RESULT 22
 ABB50275
 ID ABB50275 standard; Protein; 478 AA.
 XX
 XX ABB50275;
 AC
 XX
 XX 08-FEB-2002 (first entry)
 DT
 XX
 XX eIF-2-associated p67 ovarian tumour marker protein, SEQ ID NO:39.
 DE
 XX
 XX Ovarian tumour marker gene; human; overexpression; upregulation;
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 KW identification; serous cystadenoma; borderline serous tumour;
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
 KW immune response pathway; cell proliferation regulation; protein folding;
 KW membrane localised; secreted; therapeutic target; cytostatic;
 KW gene therapy; vaccine.
 OS
 XX Homo sapiens.
 XX
 XX WO200175177-A2.
 XX
 XX 11-OCT-2001.
 PD
 XX
 XX 03-APR-2001; 2001WO-US10947.
 PF
 XX
 XX 03-APR-2000; 2000US-194336P.
 PR
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 PI
 XX
 XX WPI; 2001-626450/72.
 DR
 XX
 XX N-P8DB; ABA83100.
 DR
 XX
 XX Detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian

PT cancer treatment, by measuring expression level of ovarian tumor marker gene -

XX

PS Claim 22; Page 87-88; 140pp; English.

XX

CC The invention relates to methods for diagnosing and prognosing ovarian tumors in an individual via the detection and measurement of the expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180, ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179, ABA83181 and ABA83183). The methods of the invention are useful for detecting an ovarian tumor in a patient, for identifying an individual at increased risk for developing ovarian cancer, in prognostic tests for assessing the relative severity of ovarian cancer, in tests for monitoring a patient in remission from ovarian cancer and in tests for monitoring disease status in a patient being treated for ovarian cancer. The methods can additionally be used to identify a particular tumor as being an ovarian tumor (i.e., an epithelial ovarian tumor selected from serous cystadenoma, borderline serous tumor, serous cystadenocarcinoma, mucinous cystadenoma, borderline mucinous tumor, mucinous cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma, clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner tumor. The ovarian tumor marker genes of the invention were identified using SAGE (serial analysis of gene expression) and were found to be overexpressed in a broad variety of ovarian epithelial tumor cells relative to normal ovarian epithelial cells. The marker genes are implicated in immune response pathways, in the regulation of cell proliferation and in protein folding, and many of these are membrane-localised or secreted. In addition to their use as diagnostic and prognostic markers, the ovarian tumor marker genes or their encoded proteins may be used as therapeutic targets for the treatment and prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent proteins encoded by ovarian tumor marker genes of the invention.

CC

XX

SQ Sequence 478 AA;

Query Match 1.9%; Score 8; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||

DB 235 NAGDTTVL 242

RESULT 23
AAB28377
ID AAB28377 standard; Protein; 478 AA.
XX
AC AAB28377;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human methionine aminopeptidase 2.
XX
XX Antinflammatory; cytosolic; antibacterial; methionine aminopeptidase 2;
XX inhibitor; MetAP2; eukaryotic initiation factor associated protein; p67;
XX eIF-2; protein synthesis; antisense oligonucleotide; infection; human;
XX inflammation; tumor.
XX
XX Homo sapiens.
XX
XX OS
XX PN US6136604-A.
XX
PD 24-OCT-2000.
XX
XX 27-OCT-1999; 99US-0428584.
XX
XX 27-OCT-1999; 99US-0428584.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Monia BP, Wyatt J;
XX

DR WPI; 2001-030942/04.
DR N-PSDB; AAC67683.
XX
XX
PT New antisense compounds which specifically hybridize with and inhibit human methionine aminopeptidase 2 expression, useful for treating methionine aminopeptidase 2 related disorders and preventing inflammation or tumor formation -
XX
XX
PS Disclosure; Columns 45-50; 39pp; English.
XX
XX
CC The present sequence is human methionine aminopeptidase 2 (also known as MetAP2 and eukaryotic initiation factor (eIF-2) associated protein, p67). MetAP2 is a cellular glycoprotein that promotes protein synthesis in the presence of active eIF-2 kinases by protecting the eIF-2 alpha subunit from phosphorylation. The present invention relates to antisense oligonucleotides (AAC67690-C67767) which inhibit expression of the present sequence. The antisense oligonucleotides of the present invention may be used for treating a patient suspected of having or being prone to a disease or condition associated with expression of MetAP2. The antisense oligonucleotides may further be used prophylactically, e.g. to prevent or delay infection, inflammation or tumor formation.

CC

XX

SQ Sequence 478 AA;

Query Match 1.9%; Score 8; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 NAGDTTVL 210
|||

DB 235 NAGDTTVL 242

RESULT 24
ABG76374
ID ABG76374 standard; protein; 478 AA.
XX
AC ABG76374;
XX
DT 20-MAY-2003 (first entry)
XX
DE Human type 2 aminopeptidase (MetAP2) variant dnwMetAP2.
XX
XX Human; dominant negative variant; type 2 methionine aminopeptidase;
XX MetAP2; translation domain; fungal infection; cell proliferation;
XX p53 function; immune system; angiogenesis; opportunistic infection;
XX cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
XX variant; dnwMetAP2.
XX
XX Homo sapiens.
XX
XX OS
XX PN Synthetic.
XX
FH Key location/Qualifiers
FT Misc-difference 219
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 231
FT /note= "Any amino acid, except His"
FT Misc-difference 251
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 262
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 328
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 331
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 338..339
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 364
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 444
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 447
FT /note= "Any naturally occurring amino acid"
FT

FT Misc-difference 459 /note= "Any naturally occurring amino acid"
 FT FT
 XX US2002182701-A1.
 PN
 XX 05-DEC-2002.
 PD
 XX 30-AUG-2001; 2001US-0943123.
 PF
 XX 30-AUG-2001; 2001US-0943123.
 PR
 XX (UYSL-) UNIV SAINT LOUIS.
 PA
 XX Chang Y, Micka WS, Vetro JA;
 PI
 XX WPI; 2003-328620/31.
 DR
 XX
 PT New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 XX
 PS Claim 3; Page 16-17; 46pp; English.
 XX
 CC The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents
 CC human MetAP2 variant dnMetAP2.
 XX
 SQ Sequence 478 AA;
 Query Match 1.9%; Score 8; DB 24; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242
 RESULT 25
 ABG76375 standard; protein; 478 AA.
 ID ABG76375
 XX
 AC ABG76375;
 XX
 DT 20-MAY-2003 (first entry)
 XX
 DE Mouse type 2 aminopeptidase (MetAP2) variant.
 XX
 KM Mouse; dominant negative variant; type 2 methionine aminopeptidase;
 KM MetAP2; translation domain; fungal infection; cell proliferation;
 KM p53 function; immune system; angiogenesis; opportunistic infection;
 KM cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
 KM variant.
 XX
 OS Mus sp.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 219 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 231 /note= "Any amino acid, except His"
 FT
 FT Misc-difference 251 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 262

FT FT /note= "Any naturally occurring amino acid"
 FT Misc-difference 328 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 331 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 338..339 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 364 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 364 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 444 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 447 /note= "Any naturally occurring amino acid"
 FT
 FT Misc-difference 459 /note= "Any naturally occurring amino acid"
 FT
 FT /note= "Any naturally occurring amino acid"
 PN US2002182701-A1.
 XX
 XX 05-DEC-2002.
 PD
 XX 30-AUG-2001; 2001US-0943123.
 PF
 XX 30-AUG-2001; 2001US-0943123.
 PR
 XX (UYSL-) UNIV SAINT LOUIS.
 PA
 XX Chang Y, Micka WS, Vetro JA;
 PI
 XX WPI; 2003-328620/31.
 DR
 XX
 PT New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 XX
 XX
 XX Claim 3; Page 17-18; 46pp; English.
 PS
 XX
 CC The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents a
 CC mouse MetAP2 variant.
 CC
 SQ Sequence 478 AA;
 Query Match 1.9%; Score 8; DB 24; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 203 NAGDTTVL 210
 |||||
 Db 235 NAGDTTVL 242
 RESULT 26
 ABG76377 standard; protein; 478 AA.
 ID ABG76377
 XX
 AC ABG76377;
 XX
 DT 20-MAY-2003 (first entry)
 XX
 DE Human type 2 aminopeptidase (MetAP2).
 XX
 KM Human; dominant negative variant; type 2 methionine aminopeptidase;
 KM MetAP2; translation domain; fungal infection; cell proliferation;
 KM p53 function; immune system; angiogenesis; opportunistic infection;
 KM cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
 KM

KW enzyme.
 XX Homo sapiens.
 OS US2002182701-A1.
 PN 05-DEC-2002.
 PD 30-AUG-2001; 2001US-0943123.
 PF 30-AUG-2001; 2001US-0943123.
 XX PR 30-AUG-2001; 2001US-0943123.
 XX (UYSL-) UNIV SAINT LOUIS.
 PA Chang Y, Micka WS, Vetro JA;
 PI WPI; 2003-328620/31.
 DR New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 CC Claim 5; Fig 1; 46pp; English.
 XX The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents human
 CC MetAP2.
 CC
 SQ Sequence 478 AA;
 QY Query Match 1.9%; Score 8; DB 24; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 203 NAGDTTVL 210
 235 NAGDTTVL 242
 RESULT 27
 ABG76378
 ID ABG76378 standard; protein; 478 AA.
 XX
 AC ABG76378;
 XX
 DT 20-MAY-2003 (first entry)
 XX
 DE Mouse type 2 aminopeptidase (MetAP2).
 XX
 KW Mouse; dominant negative variant; type 2 methionine aminopeptidase;
 KW MetAP2; translation domain; fungal infection; cell proliferation;
 KW p53 function; immune system; angiogenesis; opportunistic infection;
 KW cancer; cyostatic; fungicide; immunomodulatory; anti-angiogenic;
 KW enzyme.
 XX
 OS Mus sp.
 XX
 PN US2002182701-A1.
 PD 05-DEC-2002.
 XX
 PF 30-AUG-2001; 2001US-0943123.
 XX
 PR 30-AUG-2001; 2001US-0943123.
 XX

PA (UYSL-) UNIV SAINT LOUIS.
 XX Chang Y, Micka WS, Vetro JA;
 PI WPI; 2003-328620/31.
 DR New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 CC Disclosure; Fig 1; 46pp; English.
 XX The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents mouse
 CC MetAP2.
 CC
 SQ Sequence 478 AA;
 QY Query Match 1.9%; Score 8; DB 24; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 203 NAGDTTVL 210
 235 NAGDTTVL 242
 RESULT 28
 ABU07490
 ID ABU07490 standard; Protein; 478 AA.
 XX
 AC ABU07490;
 XX
 DT 28-JAN-2003 (first entry)
 XX
 DE Protein differentially regulated in prostate cancer #93.
 XX
 KW Prostate cancer; gene expression; differential regulation;
 KW molecular marker; drug target; cancer detection; cancer diagnosis;
 KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
 XX
 OS Homo sapiens.
 XX
 PN WO200281638-A2.
 PD 17-OCT-2002.
 XX
 PF 08-APR-2002; 2002WO-US10824.
 XX
 PR 06-APR-2001; 2001US-281731P.
 PR 06-APR-2001; 2001US-281732P.
 XX
 PA (ORIG-) ORIGENE TECHNOLOGIES INC.
 XX
 PI Sun Z, Jay G;
 XX
 DR WPI; 2003-058520/05.
 DR N-PSDB; ABX10395.
 XX
 PT Novel genes which are differentially regulated in prostate cancer,
 PT useful for diagnosing prostate cancer in prostate tissue sample and
 PT assessing therapeutic or preventive intervention in prostate cancer
 PT patients -
 XX
 PS Claim 1; Page 398-399; 416pp; English.

XX The invention describes genes (I) which are differentially regulated in prostate cancer. (I) is useful for diagnosing a prostate cancer in a sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises prostate cancer. (I) is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to prostate cancer permits the definition of functional and disease pathways and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. This is the amino acid sequence of a protein differentially regulated in prostate cancer.

XX Sequence 478 AA;

Query Match 1.9%; Score 8; DB 24; Length 478;

Best Local Similarity 100.0%; Pred. No. 41; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
|||||||
Db 235 NAGDTTVL 242

RESULT 29

ABG76381

ID ABG76381 standard; protein; 480 AA.

XX AC ABG76381;

XX DT 20-MAY-2003 (first entry)

XX DE Rat type 2 aminopeptidase (MeAP2) variant dnmMeAP2.

XX Rat; dominant negative variant; type 2 methionine aminopeptidase; MeAP2; translation domain; fungal infection; cell proliferation; p53 function; immune system; angiogenesis; opportunistic infection; cancer; cytosolic; fungicide; immunomodulatory; anti-angiogenic; variant; dnmMeAP2.

OS Ratus sp.

XX Synthetic.

XX Key Location/Qualifiers

XX FH Misc-difference 219

XX FT /note= "Any naturally occurring amino acid"

XX FT Misc-difference 231

XX FT /note= "Any amino acid, except His"

XX FT Misc-difference 251

XX FT /note= "Any naturally occurring amino acid"

FT Misc-difference 262
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 328
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 331
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 338..339
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 364
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 444
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 447
FT /note= "Any naturally occurring amino acid"
FT Misc-difference 459
FT /note= "Any naturally occurring amino acid"

PN US2002182701-A1.

XX 05-DEC-2002.

XX 30-AUG-2001; 2001US-0943123.

XX 30-AUG-2001; 2001US-0943123.

XX (USL-) UNIV SAINT LOUIS.

XX Chang Y, Micka WS, Vetro JA;

XX WPI; 2003-328620/31.

XX New variant type 2 methionine aminopeptidase polypeptide, useful for preparing a medicament for treating a disease mediated by fungal infection, cell proliferation, decreased function of p53, immune system activity or angiogenesis -

XX Claim 3; Page 28-29; 46pp; English.

XX The present invention relates to dominant negative variants of type 2 methionine aminopeptidase (MeAP2) containing a translation domain. The MeAP2 variant polypeptides are useful for preparing a medicament for treating a disease mediated by fungal infection, cell proliferation, decreased function of p53, immune system activity, CC or preferably angiogenesis. The MeAP2 polypeptides of the CC invention may be used for treating subjects suffering from cancer, CC diseases mediated by the immune system or opportunistic infections CC using inhibitors of MeAP2. The present sequence represents CC rat MeAP2 variant dnmMeAP2.

XX Sequence 480 AA;

Query Match 1.9%; Score 8; DB 24; Length 480;

Best Local Similarity 100.0%; Pred. No. 41; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 203 NAGDTTVL 210
|||||||
Db 235 NAGDTTVL 242

RESULT 30

ABG76382

ID ABG76382 standard; protein; 480 AA.

XX AC ABG76382;

XX DT 20-MAY-2003 (first entry)

XX DE Rat type 2 aminopeptidase (MeAP2).

XX Rat; dominant negative variant; type 2 methionine aminopeptidase; MeAP2; translation domain; fungal infection; cell proliferation; p53 function; immune system; angiogenesis; opportunistic infection;

KW cancer; cytostatic; fungicide; immunomodulatory; anti-angiogenic;
 KW enzyme.
 XX Rattus sp.
 OS US2002182701-A1.
 XX PN
 XX 05-DEC-2002.
 PD
 XX 30-AUG-2001; 2001US-0943123.
 PF
 XX 30-AUG-2001; 2001US-0943123.
 PR
 XX (USL-) UNIV SAINT LOUIS.
 PA
 XX Chang Y, Micka WS, Vetro JA;
 PI
 XX WPI; 2003-328620/31.
 DR
 XX
 PT New variant type 2 methionine aminopeptidase polypeptide, useful for
 PT preparing a medicament for treating a disease mediated by fungal
 PT infection, cell proliferation, decreased function of p53, immune system
 PT activity or angiogenesis -
 PS
 XX Disclosure; Fig 1; 46pp; English.
 XX
 CC The present invention relates to dominant negative variants of
 CC type 2 methionine aminopeptidase (MetAP2) containing a translation
 CC domain. The MetAP2 variant polypeptides are useful for preparing a
 CC medicament for treating a disease mediated by fungal infection, cell
 CC proliferation, decreased function of p53, immune system activity,
 CC or preferably angiogenesis. The MetAP2 polypeptides of the
 CC invention may be used for treating subjects suffering from cancer,
 CC diseases mediated by the immune system or opportunistic infections
 CC using inhibitors of MetAP2. The present sequence represents rat
 CC MetAP2.
 CC
 SO Sequence 480 AA;
 QY
 DB 203 NAGDTTVL 210
 235 NAGDTTVL 242
 |||||
 Query Match 1.9%; Score 8; DB 24; Length 480;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 31
 AAG73991
 ID AAG73991 standard; Protein; 500 AA.
 XX
 AC AAG73991;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:4755.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 PD
 XX 05-APR-2001.
 PF
 XX 28-SEP-2000; 2000WO-US26524.
 PR
 XX 29-SEP-1999; 99US-0157137.
 PR
 XX 03-NOV-1999; 99US-0163280.
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI; 2001-235357/24.
 DR N-PSDB; AAH33422.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11; Page 6553-6555; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SO Sequence 500 AA;
 QY
 DB 203 NAGDTTVL 210
 257 NAGDTTVL 264
 |||||
 Query Match 1.9%; Score 8; DB 22; Length 500;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 32
 ABG17833
 ID ABG17833 standard; Protein; 545 AA.
 XX
 AC ABG17833;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #17824.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD
 XX 11-OCT-2001.
 PF
 XX 30-MAR-2001; 2001WO-US08631.
 PR
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS82020.
 XX

XX Claim 20; SEQ ID No 60426; 103pp; English.
PS
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantifying a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 689 AA;

Query Match 1.9%; Score 8; DB 22; Length 689;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VGVVTVHP 54
|||
Db 290 VGVVTVHP 297

RESULT 35
AAU34554
ID AAU34554 standard; Protein; 891 AA.
XX
AC AAU34554;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #135.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52413.
XX

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 10147; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 891 AA;

Query Match 1.9%; Score 8; DB 22; Length 891;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RAALAAA 17
|||
Db 36 RAALAAA 43

RESULT 36
AAU38250
ID AAU38250 standard; Protein; 892 AA.
XX
AC AAU38250;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #141.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS56109.
XX

Query Match 1.9%; Score 8; DB 22; Length 1300;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RAALALAA 17
 |||||
 DB 877 RAALALAA 884

RESULT 39

ABG05084
 ID ABG05084 standard; Protein; 1882 AA.

XX ABG05084;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5075.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS69271.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID NO 35443; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1882 AA;

Query Match 1.9%; Score 8; DB 22; Length 1882;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 QIABGKLL 243
 |||||
 DB 1729 QIABGKLL 1736

RESULT 40

AAM40183
 ID AAM40183 standard; Protein; 1883 AA.

XX AAM40183;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3328.

KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0596042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;

DR WPI: 2001-442253/47.

DR N-PSDB; AAI59339.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

PS Example 5; SEQ ID NO 3328; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM18642-AAM42213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX

SQ Sequence 1883 AA;

Query Match 1.9%; Score 8; DB 22; Length 1883;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 QIAEGKLL 243

|||||

Db 1730 QIAEGKLL 1737

Search completed: September 8, 2003, 14:08:37
 Job time : 62 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:56:26 ; Search time 41 Seconds
(without alignments)
2593.114 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019 1 MAFYAFKRAMRAALAAVAL.....AAPQSGVQTASEAKTASEAR 412

Sequence:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Scoring table:

Searched: 830525 seqs, 258052604 residues

830525

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp Unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	1993	98.7	412	16	Q9J50 neisseria m
2	1989	98.5	412	16	Q9J66 neisseria m
3	1278	63.3	271	2	Q51007 neisseria g
4	836	41.4	398	16	Q8Y39 raltosonia s
5	777	38.5	385	16	Q8ZLN5 salmonella s
6	770	38.1	388	16	Q9AA03 caulobacter
7	765	37.9	399	16	Q8P7C8 xanthomonas
8	763.5	37.8	399	2	Q9ABG2 enterobacte
9	760.5	37.7	409	16	Q8CWA9 escherichia
10	758	37.5	409	16	Q8XRL2 raltosonia s
11	756	37.4	397	16	Q8ZBT7 salmonella
12	753	37.3	397	16	Q8ZRA6 salmonella
13	752.5	37.3	396	2	Q8VPA9 proteus mir
14	749.5	37.1	385	16	Q8X4W2 escherichia
15	745.5	36.9	385	16	Q8CVN7 escherichia
16	743.5	36.8	394	2	Q9F241 xanthomonas

17	741.5	36.7	404	16	Q8PIQ1 xanthomonas
18	734	36.4	382	16	Q8EBH1 shewanella
19	730.5	36.2	384	2	Q9WZ9 pseudomonas
20	727.5	36.0	408	16	Q8PJ6 xanthomonas
21	726.5	36.0	395	16	Q8ZC88 yersinia pe
22	723	35.8	427	16	Q8UB14 agrobacteri
23	720	35.7	371	2	Q9KUC3 pseudomonas
24	719.5	35.6	407	16	Q9ZT02 rhizobium m
25	719	35.6	374	2	Q8GC84 enterobacte
26	701.5	34.7	384	16	Q8CW42 escherichia
27	700.5	34.7	388	2	Q8RSM2 uncultured
28	699.5	34.6	387	16	Q5I395 pseudomonas
29	697.5	34.5	373	16	Q8X7E1 escherichia
30	683	33.8	388	2	Q93K41 klebsiella
31	670	33.2	408	16	Q9BPB7 xylella fas
32	662	32.8	398	2	Q9RBY9 xanthomonas
33	655	32.4	385	16	Q8X4L0 escherichia
34	655	32.4	385	16	Q8CVL1 escherichia
35	654	32.4	382	2	Q9KWV5 pseudomonas
36	647.5	32.1	436	16	Q8P876 xanthomonas
37	647	32.0	396	2	Q93E20 acinetobact
38	640	31.7	382	2	Q3I099 pseudomonas
39	636	31.5	435	16	Q8P8N2 xanthomonas
40	631	31.3	391	2	Q93P05 pseudomonas
41	628.5	31.1	386	16	Q8ZBD3 yersinia pe
42	628	31.1	396	16	Q9RG60 pseudomonas
43	607.5	30.1	399	2	Q87935 burkholderi
44	606	30.0	385	16	Q8XT04 raltosonia s
45	602	29.8	384	2	Q68440 agrobacteri

ALIGNMENTS

RESULT 1

Q9J50 PRELIMINARY; PRT; 412 AA.

AC Q9J50; (TREMREL1, 15, Created)

DT 01-OCT-2000 (TREMREL1, 15, Last sequence update)

DT 01-OCT-2000 (TREMREL1, 22, Last annotation update)

DE Membrane fusion protein.

GN MTRC OR NMA1970.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

OX Neisseriaceae; Neisseria.

OX NCBI_TaxID=5699;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Z2491 / Serogroup A / Serotype 4A;

RX MEDLINE=40222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S., Jagers K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G., "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."

RT Nature 404:502-506(2000).

RL EMBL; AL162757; CAB85190.1; -

DR InterPro; IPR006143; HlyD.

DR Pfam; PF00529; HlyD; 1.

KW Complete proteome.

SO SSOURCE 412 AA; 42954 MW; SCF797BA370AA75D CRC64;

Query Match 98.7%; Score 1993; DB 16; Length 412;

Best Local Similarity 98.5%; Score No. 5,8e-102; Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAFYAFKRAMRAALAAVALVSSCGKGGDAQGGGPAGREAPAPVGVTVHPQVALT 60

DB 1 MAFYAFKRAMRAALAAVALVSSCGKGGDAQGGGPAGREAPAPVGVTVHPQVALT 60

```

QY 61 VELPGRLSRLTADVRAOVGGIIQKRLFOEGSYVRAQGPLYQIDSTYEANLESARAQLA 120
Db 61 VELPGRLSRLTADVRAOVGGIIQKRLFOEGSYVRAQGPLYQIDSTYEANLESARAQLA 120
QY 121 TAQATLAKADADLARYPFLVAEAVSRQEDAAVTAKRSABAGVKAQAIAIKSAGINLNR 180
Db 121 TAQATLAKADADLARYPFLVAEAVSRQEDAAVTAKRSABAGVKAQAIAIKSAGINLNR 180
QY 181 SRITAPISGFGQSKVSEGLTLNAGDTTVLATTIRQTPMVMYNTQSAEVMKLRQIAEG 240
Db 181 SRITAPISGFGQSKVSEGLTLNAGDTTVLATTIRQTPMVMYNTQSAEVMKLRQIAEG 240
QY 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAAVPNDONILMPGLYV 300
Db 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAAVPNDONILMPGLYV 300
QY 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIWTSGLKD 360
Db 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIWTSGLKD 360
QY 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412
Db 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412

RESULT 2
ID Q9JY66 PRELIMINARY; PRT; 412 AA.
AC Q9JY66;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Membrane fusion protein.
GN NMB1716.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
CX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tectelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Halt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Cotton M.D., Ueberback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiarini V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunoi R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002521; AAF42063.1; -.
DR TIGR; NMB1716; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 42795 MW; 0B7A0DB0E3113E4E CRC64;

Query Match 98.5%; Score 1989; DB 16; Length 412;
Best Local Similarity 98.5%; Pred. No. 9.7e-102;
Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 121 TAQATLAKADADLARYPFLVAEAVSRQEDAAVTAKRSABAGVKAQAIAIKSAGINLNR 180
Db 121 TAQATLAKADADLARYPFLVAEAVSRQEDAAVTAKRSABAGVKAQAIAIKSAGINLNR 180
QY 181 SRITAPISGFGQSKVSEGLTLNAGDTTVLATTIRQTPMVMYNTQSAEVMKLRQIAEG 240
Db 181 SRITAPISGFGQSKVSEGLTLNAGDTTVLATTIRQTPMVMYNTQSAEVMKLRQIAEG 240
QY 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAAVPNDONILMPGLYV 300
Db 241 KLLAADGVIAVGIKFDGTVYPEKGRLLPADPVNVESTGOITLRAAVPNDONILMPGLYV 300
QY 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIWTSGLKD 360
Db 301 RVLMDQVAVDNAFVVPQAAVTRGAKDTVMIVNAQGMPEPREVTVAQOQGTNMIWTSGLKD 360
QY 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412
Db 361 GDKVVEGISIAGITGAKKVTPEKMASSENOAAAPQSGVOTASAKTASEAE 412

RESULT 3
ID Q51007 PRELIMINARY; PRT; 271 AA.
AC Q51007;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE MtrC protein (Fragment).
GN MTRC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
CX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH95;
RX MEDLINE=94254732; PubMed=8196548;
RA Pan W., Spratt B.G.;
RT "Regulation of the permeability of the gonococcal cell envelope by the
RT mtr system."
RL Mol. Microbiol. 11:769-775 (1994).
DR EMBL; Z25796; CA81046.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW NON_TER
SQ SEQUENCE 271 AA; 28145 MW; 1D0ADD1F335B39C5 CRC64;

Query Match 63.3%; Score 1278; DB 2; Length 271;
Best Local Similarity 96.7%; Pred. No. 6.6e-63;
Matches 262; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```


RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohn N., Maddock J.R.,
 RA Porocka J., Nelson W.C., Newton M.C., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Dirkin A.S., Gwinn M.L., Half D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khoult H., Shetty J., Berry K.,
 RA Utecher T., Tran K., Wolf A., Yamathavan J., Ernoelava M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RA "Complete genome sequence of *Caulobacter crescentus*,"
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RL EMBL; AF005757; AAK22793.1; -.
 DR TIGR; CC0808; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 KW Complete proteome.
 SQ SEQUENCE 388 AA; 39483 MW; 40934C3F69052DAD CRC64;

Query Match 38.1%; Score 770; DB 16; Length 388;
 Best Local Similarity 46.2%; Pred. No. 8.9e-35;
 Matches 180; Conservative 57; Mismatches 131; Indels 22; Gaps 9;

16 AAVLVVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 74
 Db AAVLVVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 74
 15 AAVLVVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 69
 Qy AAVLVVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 69
 Db VRAOVGIIQKRLFOEGSVYRAGOPLYOIDSSTYANLESARAPL-ATAQATLAKAD 132
 75 VRAOVGIIQKRLFOEGSVYRAGOPLYOIDSSTYANLESARAPL-ATAQATLAKAD 132
 Db VRPOVSGVIOKRLPEEGAVNRAQOPLYOIDPATYAAANSAALAAQOATLAKAD 129
 70 VRPOVSGVIOKRLPEEGAVNRAQOPLYOIDPATYAAANSAALAAQOATLAKAD 129
 Qy 133 LARYPLVAEAVSRQEDYAAVTAKRSAEAGVKAQAIAKAGINLNSRTAPISFIG 192
 133 LARYPLVAEAVSRQEDYAAVTAKRSAEAGVKAQAIAKAGINLNSRTAPISFIG 192
 Db 130 -RYVALVETGAVSRQEDYAAVTAKRSAEAGVKAQAIAKAGINLNSRTAPISFIG 187
 130 -RYVALVETGAVSRQEDYAAVTAKRSAEAGVKAQAIAKAGINLNSRTAPISFIG 187
 Qy 193 OSKSYSEGLLNAAGDTTLATIRQTNPMYVNTQSGSEFMKLROIAEKLAAQVAVG 252
 193 OSKSYSEGLLNAAGDTTLATIRQTNPMYVNTQSGSEFMKLROIAEKLAAQVAVG 252
 Db 188 KSSVYAGLVANQATLALATVQDSKVVVDLTQTSABLLKLOAQPSAGK-VGRSGSAQVT 246
 188 KSSVYAGLVANQATLALATVQDSKVVVDLTQTSABLLKLOAQPSAGK-VGRSGSAQVT 246
 Qy 253 IKFDGTVYPERGRLLFADPVVNESTQITLRAAVPNDQNLMPGLYRVLMQVAVDNA 312
 253 IKFDGTVYPERGRLLFADPVVNESTQITLRAAVPNDQNLMPGLYRVLMQVAVDNA 312
 Db 247 LKLBDSGYPIPRGRLEFSDITVDPTGAVGLRAVDNKGKVLIPMYRAVAVLSQVAVASG 306
 247 LKLBDSGYPIPRGRLEFSDITVDPTGAVGLRAVDNKGKVLIPMYRAVAVLSQVAVASG 306
 Qy 313 FVDPQAVTRAKD-TWIVNAQGMREPVYVAAQOGTMMVITSGKDDKVVVGIS 370
 313 FVDPQAVTRAKD-TWIVNAQGMREPVYVAAQOGTMMVITSGKDDKVVVGIS 370
 Db 307 ILIIPOTAVNRDPKGGATVLMVGAK-GPEPRPVTLIGQTVGDKWLVTSGLNAGDKVIEGL- 364
 307 ILIIPOTAVNRDPKGGATVLMVGAK-GPEPRPVTLIGQTVGDKWLVTSGLNAGDKVIEGL- 364
 Qy 371 IAGITGAKKVTPEKMASSENQAAAPQSGVQ 400
 371 IAGITGAKKVTPEKMASSENQAAAPQSGVQ 400
 Db 365 -----MKVRPGAPIKAVPAGAAPAAQAQ 387
 365 -----MKVRPGAPIKAVPAGAAPAAQAQ 387

RESULT 7

08P7C8 PRELIMINARY; PRT; 399 AA.
 AC 08P7C8;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Multidrug resistance protein.
 GN MEYA OR KCC2683.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 CX NCB1_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCEPB 528;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Ousaglo R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chamebo J.F., Clapina L.P.,
 RA Ciccarelli R.M.B., Coutinho L.D., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira L.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Terubal J.C., Kitajima J.P.,
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities,"
 RT Nature 417:459-463(2002).
 RL EMBL; AE012380; AA041955.1; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 41511 MW; 59C1197DE605D2D CRC64;

Query Match 37.9%; Score 765; DB 16; Length 399;
 Best Local Similarity 43.3%; Pred. No. 1.7e-34;
 Matches 174; Conservative 67; Mismatches 139; Indels 22; Gaps 6;

8 AMRAALAAVALVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 65
 Db 8 AMRAALAAVALVSSCG-KGSDAAGGPPAGREAPVGVVTHPQVTLTVLPRLSLRTAD 65
 6 SLRLPLVAASVLSACGSPGPPQG-----TPVGVITVRPOPVTTLTQLPG 57
 Qy 6 SLRLPLVAASVLSACGSPGPPQG-----TPVGVITVRPOPVTTLTQLPG 57
 Db RLSLRTADRAOVGIIQKRLFOEGSVYRAGOPLYOIDSSTYANLESARAPLATAQAT 125
 66 RLSLRTADRAOVGIIQKRLFOEGSVYRAGOPLYOIDSSTYANLESARAPLATAQAT 125
 Db 58 RTVYVLAERAPQGVIGVQTRQFEGSDVAKAGOTLYOIDPATIRASVAAQATLAKAD 117
 58 RTVYVLAERAPQGVIGVQTRQFEGSDVAKAGOTLYOIDPATIRASVAAQATLAKAD 117
 Qy 126 LAKADADLARYKPLVAEAVSRQEDYAAVTAKRSAEAGVKAQAIAKAGINLNSRTAP 185
 126 LAKADADLARYKPLVAEAVSRQEDYAAVTAKRSAEAGVKAQAIAKAGINLNSRTAP 185
 Db 118 LRTARLADRYKELVQKALISQEGDDTAATLQGAADVAAGKASVETARINLAFAMD 177
 118 LRTARLADRYKELVQKALISQEGDDTAATLQGAADVAAGKASVETARINLAFAMD 177
 Qy 186 PISGFIQSKVSEGLLNAAGDTTLATIRQTNPMYVNTQSGSEFMKLROIAEKLAA 244
 186 PISGFIQSKVSEGLLNAAGDTTLATIRQTNPMYVNTQSGSEFMKLROIAEKLAA 244
 Db 178 PISGRIRSSVTPALVTAQATLTTIQJLDPIYDVTPSAVAVLKKAMAGDLARA 237
 178 PISGRIRSSVTPALVTAQATLTTIQJLDPIYDVTPSAVAVLKKAMAGDLARA 237
 Qy 245 ADGVIAVGIFKFDGTVYPERGRLLFADPVVNESTQITLRAAVPNDQNLMPGLYRVLM 304
 245 ADGVIAVGIFKFDGTVYPERGRLLFADPVVNESTQITLRAAVPNDQNLMPGLYRVLM 304
 Db 238 GDGAQVSVLVEDOSTYFLOGRLAFSDVTVQNTGSLRAVFPNPADLIPGYVAVL 297
 238 GDGAQVSVLVEDOSTYFLOGRLAFSDVTVQNTGSLRAVFPNPADLIPGYVAVL 297
 Qy 305 DQVAVDNAFVPPQAVTR--GAKDTVMVNAQGMREPVYVAAQOGTMMVITSGKDD 362
 305 DQVAVDNAFVPPQAVTR--GAKDTVMVNAQGMREPVYVAAQOGTMMVITSGKDD 362
 Db 298 QEGVKAQGVLPQAVTRNAGKPTAFVGAADNKLQRLVETBRAVDQWLVRSGLKTGD 357
 298 QEGVKAQGVLPQAVTRNAGKPTAFVGAADNKLQRLVETBRAVDQWLVRSGLKTGD 357
 Qy 363 KVVVEGISTAGITGAKKVTPEK-----ASSNQAAAPQSGVQ 397
 363 KVVVEGISTAGITGAKKVTPEK-----ASSNQAAAPQSGVQ 397
 Db 358 QLVVDGLSRARDGVQVKTVP--WQPTSASVAGSPSAPAAPRA 397
 358 QLVVDGLSRARDGVQVKTVP--WQPTSASVAGSPSAPAAPRA 397

RESULT 8

09AEG2 PRELIMINARY; PRT; 399 AA.
 AC 09AEG2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Acra protein precursor.
 GN ACRA.
 OS Enterobacter aerogenes (Aerobacter aerogenes).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Enterobacter.
 CX NCB1_TaxID=548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BW16627;
 RX Pradel B., Pages J.M.,
 RA "The *Acra*/*AcraB*/*tolC* efflux pump participates in multidrug resistance
 RT in *Enterobacter aerogenes*,"
 RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ306389; CA035724.1; -.
 DR InterPro; IPR006143; HlyD.
 DR Pfam; PF00529; HlyD; 1.

KM Signal. 1 24 POTENTIAL.
 FT SIGNAL 25 399 ACRA.
 FT CHAIN 399 AA; 42443 MM; AC49F8E3870B6E78 CRC64;
 SQ SEQUENCE

Query Match 37.8%; Score 763.5; DB 2; Length 399;
 Best Local Similarity 42.0%; Pred. No. 2,1e-34;
 Matches 169; Conservative 80; Mismatches 134; Indels 19; Gaps 5;

QY 11 AALAAALVALVSSCGGDAAGGAPAGREAPVAVVTVHPQTALVTELGRLSL 70
 DB 11 AVTLMSSSLALVTCG-DKPKAQQGAAQ-----QMPVGVITLKSAPIQITTELPRTNAY 63
 QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARQATATQATLAKAD 130
 DB 64 RVAEVRPQVSGIILKRNFTGESSDIOAGVSLYQIDPATYQAYESAKGDLAKQAQAAANIAQ 123
 QY 131 ADLARYKPLVAAEAVSROEYDAVATKRSABAGYKAAQAALKSAGININRSRTAPISGF 190
 DB 124 LTVRRYQKLTGYTKYSIOEYDASAVADAQOSNAVAVAKAAVETARINILATYKVTSPISGR 183
 QY 191 IGSQKSEGTLLNAGDTTVALTITQTPMNVNTQSSSEWKLRLQIAEGKLAADGVIA 250
 DB 184 IGKSAVTEGALVONGQSTALATVQGLDPIYDVQSSNDPLRLKQELANGTLKQENGKAK 243
 QY 251 VGIKFDGTAVYPEKGRLLFADPVVNESTGQITLRAAVPNQNIIMPGLYAVLMDQVAVD 310
 DB 244 VELVTNGLKAKYPOGDTLEFSDVTVDTGTSITLRAIFPNPDHLLPGMFVRAAREBGINP 303
 QY 311 NAFVVPQQAATVRGAKD--TWMTVNAQGMREPVTVAAQOQGTWNI VTSGLKDGKRVVEG 368
 DB 304 DALVLPQOGVTRTPRGADATVAVVEGDEVEVRQVTAQAIDDKMLVTEGLKSGDRVIYAG 363
 QY 369 ISAGITGAKKVTPEKMASSSENQAAAPQSGVQTAASEAKTASE 410
 DB 364 L-----QKVRGVQVKAQEVYADNQQQ--QTAGNNAQOSE 395

RESULT 9

08CWA9 PRELIMINARY; PRT; 409 AA.
 AC 08CWA9;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Acriflavine resistance protein A precursor.
 GN ACRA OR C0581.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burdand V., Plunkett G. III, Redford P., Roesch P.,
 RA Raske D., Buckles E.L., Biou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.P., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AEO16756; AAN79059.1; -.
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 43521 MM; 4234179F03G60197 CRC64;

Query Match 37.7%; Score 760.5; DB 16; Length 409;
 Best Local Similarity 42.2%; Pred. No. 3,2e-34;
 Matches 166; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

QY 11 AALAAALVALVSSCGGDAAGGAPAGREAPVAVVTVHPQTALVTELGRLSL 70
 DB 23 AVTLMSSSLALVTCG-DKPKAQQGAAQ-----QMPVGVITLKSAPIQITTELPRTNAY 75

QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARQATATQATLAKAD 130
 DB 76 RVAEVRPQVSGIILKRNFTGESSDIOAGVSLYQIDPATYQAYESAKGDLAKQAQAAANIAQ 135
 QY 131 ADLARYKPLVAAEAVSROEYDAVATKRSABAGYKAAQAALKSAGININRSRTAPISGF 190
 DB 136 LTVRRYQKLTGYTKYSIOEYDADQALADAQOANNAVAATAKAAVETARINILATYKVTSPISGR 195
 QY 191 IGSQKSEGTLLNAGDTTVALTITQTPMNVNTQSSSEWKLRLQIAEGKLAADGVIA 250
 DB 196 IGKSNVTEGALVONGQATATATVQGLDPIYDVQSSNDPLRLKQELANGTLKQENGKAK 255
 QY 251 VGIKFDGTAVYPEKGRLLFADPVVNESTGQITLRAAVPNQNIIMPGLYAVLMDQVAVD 310
 DB 256 VSLTSPGKIFPOGDTLEFSDVTVDTGTSITLRAIFPNPDHLLPGMFVRAAREBGINP 315
 QY 311 NAFVVPQQAATVRGAKD--TWMTVNAQGMREPVTVAAQOQGTWNI VTSGLKDGKRVVEG 368
 DB 316 NALVLPQOGVTRTPRGADATVAVVEGDEVEVRQVTAQAIDDKMLVTEGLKSGDRVIYAG 375
 QY 369 ISAGITGAKKVTPEKMASSSENQAAAPQSGVQTAASEAKTASE 400
 DB 376 L-----QKVRGVQVKAQEVYADNQQQAA--SGAQ 403

RESULT 10

08XRL2 PRELIMINARY; PRT; 409 AA.
 AC 08XRL2;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Probable drug efflux lipoprotein.
 GN RSP0819 OR RS0188.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Ralstoniaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GMI1000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Selaoubt M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Broctier P., Camus J.C., Catalicio B.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RL Nature 415:497-502(2002).
 DR EMBL; AL646081; CAD17970.1; -.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 409 AA; 42873 MM; 97003B2B7F95401B CRC64;

Query Match 37.5%; Score 758; DB 16; Length 409;
 Best Local Similarity 44.4%; Pred. No. 4,4e-34;
 Matches 182; Conservative 68; Mismatches 136; Indels 24; Gaps 9;

QY 12 AALAAALVALVSSCGGDAAGGAPAGREAPVAVVTVHPQTALVTELGRLSL 68
 DB 13 AALAA--ASLLAACGK-----PPGGPPPEAGCTPVVGVMTVQPORVTLDTLDELGRTV 61
 QY 69 SLRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARQATATQATLAK 128
 DB 62 PLVAVPQVQVNGIILKRNFTGESSDIOAGVSLYQIDPATYQAYESAKGDLAKQAQAAANIAQ 121
 QY 129 ADLARYKPLVAAEAVSROEYDAVATKRSABAGYKAAQAALKSAGININRSRTAPIS 188
 DB 122 TRIKARHKEIVALIQAIVSRQDYDDAAALAGEADVAARAVETSRINILAVARVDAPIS 181

Oy		189	GFIGOSKXSEGLTNAGDTVTATIRONTNPMWVNTQGSASEVMKR.RROIAEKKLLAAGCV	248
Dd		182	GRIGSSVTPGALVPANOTSTIATTQQDLPYVDVTQPSAALLRURQAMARDELQKSGAN	241
Oy		249	IA-VAIKFDDGVVPEKGRLLFPADVNVVESTGOITLRAAPNDONILMPGLYVRVLMDOV	307
Dd		242	AATVRLLLEDGSAFLVEGLEFRSDVTVOGNTSCTLRAVFPRPSPAILLPGMYRVRAVBEG	301
Oy		308	AVDNAFVVYPQAATVGA--KDTMTLVNAOAGMEPEBVTVAAQOOGTNNIVTSGLKDGVY	365
Dd		302	VDEALFLVPOQAARDSTGCKPPAYVVGNDRKLRRTLETERFTVGOWMVRSGSLRGIDOLV	361
Oy		366	VEGISIIAGITGAK-KVTPKEW---ASSENQAAAPSOVSQOTASEATASRA	411
Dd		362	VEGLPRA-VPGAENVKTP--WTGKTATISNPAAAPPAVTAAGORTVAA	408
<hr/>				
	RESULT	11		
ID	0828T7	PRELIMINARY;	PRT:	397 AA.
AC	0828T7			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, last annotation update)			
DE	Acridavin resistance protein A.			
CN	STR0520.			
OS	Salmomella typhi.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
CC	Enterobacteriaceae; Salmomella.			
OX	NCBI_TaxID=601;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CT18;			
EX	MEDLINE=21534947; PubMed=11677608;			
RA	Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,			
RA	Croun A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagals K.,			
RA	Krogh A., Larsen T.S., Leachner S., Moulie S., O'Gaora P., Parry C.,			
RA	Quell M., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RT	"Complete genome sequence of a multiple drug resistant Salmomella			
RL	enterica serovar Typhi CT18.";			
RT	Nature 413:848-852(2001).			
DR	EMBL; AL627267; CAD04961.1; --			
DR	InferPro; IPR006143; HYD.			
DR	Pfam; PF00529; HYD; 1.			
KV	Complete proteome.			
SC	SEQUENCE 397 AA; 42221 MW; F667D42FD8B9EEBD CRC64;			
<hr/>				
	Query Match	37.4%;	Score 756;	DB 16; Length 397;
	Best Local Similarity	41.3%;	Pred. No. 5,4e-34;	
	Matches 166; Conservative	78;	Mismatches 128;	Indels 30; Gaps 5
Oy		11	AAALAANAALVLSGCGKGDAAGQGPARGREAPRVNGVTHPCQTVALTYVELPERLESL	70
Dd		11	AVLTMLSSGLALTGCDDKDD-QQGQ-----QMPEVGVTILKTEPLPDITTELPERTVAY	63
Oy		71	RTADVRAOVGGIIOKRLFQESGYVRAGOPLYOIDSYTEANLESARAQDLATAQATLAD	130
Dd		64	RIAEVRPVQSGIILKRNFEVGSIDIEAGVSLVQIDPATVQATYDSAKGDLAKAQAAANIAB	123
Oy		131	ADLARPKVLVAEAENSROETPYDAATVATKASALEGVYAQAQAATKSAGINLNRSKITPI	190
Dd		124	LTVKRYQQLTGTOYISKQBYDALADAQAQATAVAANAANAETAANIATLYTKVTSPISGR	183
Oy		191	IGOSKVSSEGTLLNAGDTVTATIRONTNPMWVNTQGSASEVMKR.RROIAEKKLLAADGYIA	250
Dd		184	IGKSSVTGALVONGQASALATVQDLDPIYDVDTOSSNDPFLRKKELANGLANGSLKQENGAK	243
Oy		251	VGIRKDDGVVPEKGRLLFPADVNVVESTGOITLRAAPNDONILMPGLYVRVLMDOVA	310

```

Db      244 VDLVTSDDIKPPQSGSTLEFSVDVTDQSGSTILRLIFPNPHDTLLPGMFVRRLRLOEGTRP 303
Qy      311 NAFVVPQQAIVTRGAKD--TWMI VNAAGGEMEREVVAOQOGSTNMIVTSGLDKDGDKVVEG 368
Db      304 TALLVPOQGVTRTRPRGDATVTVVGADNKNVEERQIVASQAIDDKMLVTTDGLKAGDRAVVVSG 363
Qy      369 ISIAGITGAKKVT-----KEMASSENQAAAPQS 397
Db      364 L-----QKVRPGAQVKEVQBITADNKKQAASGDPAQPRS 397

RESULT 12
O8ZRA6
ID      O8ZRA6      PRELIMINARY;      PRT;      397 AA.
AC      O8ZRA6;
DT      01-MAR-2002 (TREMBLrel. 20, Created)
DT      01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT      01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE      Acridine efflux pump.
ACRA OR STM0476.
OS      Salmonella typhimurium.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Salmonella.
OC      NCBI_Taxid:602;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX      MEDLINE=21534948; PubMed=11677609;
RA      McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA      Courtney L., Porwollik S., Aili J., Dante M., Du F., Hou S., Layman D.,
RA      Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA      Waterston R., Wilson R.K.;
RT      "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT      LT2."
RIL      Nature 413:852-856(2001).
DR      EMBL; AE008717; AAL19430.1; -.
DR      InterPro; IPR006143; H1YD.
DR      Pfam; PF00529; H1YD.
KW      Complete proteome.
SQ      SEQUENCE 397 AA; 42235 MW; E363C42FDB8BDEBE CRC64;

Query Match      37.3%; Score 753; DB 16; Length 397;
Best Local Similarity 41.0%; Pred. No. 7.9e-34;
Matches 165; Conservative 79; Mismatches 128; Indels 30; Gaps 5;

11 AALAAAVLVLSGCGKGGDAAGCGQAPRGRAAPRVGVVTVHQTVALTYELPRLESL 70
11 AVVLMLSGLSLALTCGDDKQD-QQGQ-----QMPEGVVTLKTEPLDITTELPRTVAY 63
71 RTADVRAOVGGIIQKRLFOEGSVYRAGQPLQIDNSTYEANLLESHRAQLATAQATLAKAD 130
64 RIAERPRPVSGLILKKNFVEGSDIEAGSLYQIDPATQATYDASKGLAKQAQAAMATAE 123
131 ADLARYPELVAAEAVSRQGYDAVATYAKRSAGVKKAAQAAIKASGININRSRTIAPISGF 190
124 LTVKRYQKLTGTQYISKQYDOALADAAQATAAVVAARAAAVETARINLATYKTPSISGR 183
191 IGGSVRSAGCTILNAGDTTVLATITQTNMYNNVNTQSASEWKLKRLQIAEGTLLAADGYIA 250
184 IGSKSVTGAIVONGQASALATVQOLDPIYDVDTQSSNDPLRLKQELNGLSKOENGKAK 243
251 VGIKRDEDTVYPEKGRILFADPVVNESVGTTLRAAVNDONILMPGLVYRVLMDOVAVD 310
244 VDLVTSDDIKPPQSGSTLEFSVDVTDQSGSTILRLIFPNPHDTLLPGMFVRRLRLOEGTRP 303
311 NAFVVPQQAIVTRGAKD--TWMI VNAAGGEMEREVVAOQOGSTNMIVTSGLDKDGDKVVEG 368
304 TALLVPOQGVTRTRPRGDATVTVVGADNKNVEERQIVASQAIDDKMLVTTDGLKAGDRAVVVSG 363
369 ISIAGITGAKKVT-----KEMASSENQAAAPQS 397

```

DB 364 L-----QKVRPGAQYKVOEITADNKNQAASGDPQAQPRS 397

RESULT 13

Q8VPA9 PRELIMINARY; PRT; 396 AA.

AC Q8VPA9: 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)

DE Membrane fusion protein Acra.

OS Proteus mirabilis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

CC Enterobacteriaceae; Proteus.

OX NCBI_TaxID=584;

RN [1]

RP SEQUENCE FROM N.A.

RA Visselli M.A., Murphy E., Projan S.J., Bradford P.A.;

RT "Acetate efflux homolog is associated with reduced levels of

RT susceptibility to GAR-936 in Proteus mirabilis."

RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AY061647; AAL32125.1; -

DR InterPro; IPR006143; HlyD.

DR Pfam; PF00529; HlyD; 1.

SO SEQUENCE 396 AA; 42256 MW; 261F3865A4746A23 CRC64;

Query Match 37.3%; Score 752.5; DB 2; Length 396;

Best Local Similarity 42.6%; Pred. No. 8.4e-34;

Matches 168; Conservative 71; Mismatches 138; Indels 17; Gaps 6;

QY 12 AALAAVALVSSCGKGDAAAGGQAPAGRAPVGVVTHPQTALTVELPGRLESIR 71

DB 11 ALLVLSGSLVAVAGCGDKNQSA-GGPP-----PAPAVGVTLDAKPLRTITDLPGRISAYR 64

QY 72 TADVRAQVGGIIOKRLFOGSSYVRAGOPLYOIDSSTYANLBSARQATQATLAKDA 131

DB 65 IAEVRPQVGGIILKKNYSSVEGSLYQIDPITFATINSAGADLAKKAETARL 124

QY 132 DLAVYKPLVAEAVSRQEDAAVTAKRSAEAGVKAQAQAIKSAGINLRSRTAPISGFI 191

DB 125 TVERYKPLGNTYNSKQDFRTISQYQAVAAVRAAEITVNAKINLETYKTAIPISGRS 184

QY 192 GQSKVSEGTLLNAGDTVLATIRQTNPMYVNVTOSSASEVMKLRQIAEGKLLAADGVIAV 251

DB 185 GKSTVTEGALVAPGQVVALTTVOQIDPIYVDVTOSSSEYDLKLNKEIESGIIROEGKRVV 244

QY 252 GIKEDDGVYPEKGRLLFADPVNVESTQITLRAAVPDQNTILMGLYVRVLMQVAVDN 311

DB 245 HLTJNSQSYAQKHLEFSDVTDESTSITMRAIVPMPKCELLPGMFVRKLENGIRON 304

QY 312 AFVVPQOAVTRGAKD--TWMIIVNAQGMPEPREVTVAQOQGTNMTVTSGLKXGDKVVVSGI 369

DB 305 AILPQOAVITPRGEATITVNNKDNVVEVRTIEVSQAVGNKMLVNSGVQGDVIVSGL 364

QY 370 SIAGITGAKKVTPE-----WASSENQAA--PQ 396

DB 365 OKA--QPEMKVTPOEENLDATASTEKSEPPADPQ 396

RESULT 14

Q8X4W2 PRELIMINARY; PRT; 385 AA.

AC Q8X4W2: 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)

DE Transmembrane protein affects septum formation and cell membrane

DE permeability.

GN ACRE OR Z4625 OR ECG4137.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

CC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"

RL Nature 409:529-533(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Haysashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsuda E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasekawa C., Ogawara N., Yasunaga T.,

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

DR EMBL; AE005354; AAG58393.1; -

DR EMBL; AP002564; BAB37560.1; -

DR InterPro; IPR006143; HlyD.

DR Pfam; PF00529; HlyD; 1.

KW Complete proteome.

SO SEQUENCE 385 AA; 41345 MW; 9ABF82C84AF2F46B CRC64;

Query Match 37.1%; Score 749.5; DB 16; Length 385;

Best Local Similarity 42.1%; Pred. No. 1.2e-33;

Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

QY 17 AVALVSSCGKGDAAAGGQAPAGRAPVGVVTHPQTALTVELPGRLESIR 71

DB 17 SAALIAQCNDDGEEKAVHGEPO-----ATVHIKTAPELVKTELPGRITAVRIA 65

QY 74 DVRAQVGGIIOKRLFOGSSYVRAGOPLYOIDSSTYANLBSARQATQATLAKDA 133

DB 66 EVRPQVGGIILKKNYSSVEGSLYQIDPITFATINSAGADLAKKAETARL 125

QY 134 ARYKPLVAEAVSRQEDAAVTAKRSAEAGVKAQAQAIKSAGINLRSRTAPISGFI 193

DB 126 KRYVPLVGTXYISQEDVQADARQDAVTAIKAIVESARINLATAKTAIPISGRGX 185

QY 194 SKVSEGTLLNAGDTVLATIRQTNPMYVNVTOSSASEVMKLRQIAEGKLLAADGVIAV 253

DB 186 STVTEGALVNVGQTTTELATVQQLDPIYVDVTOSSSEYDLKLNKEIESGIIROEGKRVV 244

QY 254 KFDGTVYPEKGRLLFADPVNVESTQITLRAAVPDQNTILMGLYVRVLMQVAVDN 313

DB 246 VMNGQVTPYKGLGTFSDVTVDESTSITLRAVFPNPOHTLPGMFVRKLENGIRON 305

QY 314 VVPQOAVTRGAKD--TWMIIVNAQGMPEPREVTVAQOQGTNMTVTSGLKXGDKVVVSGI 371

DB 306 LIPQOAVITPRGEATITVNNKDNVVEVRTIEVSQAVGNKMLVNSGVQGDVIVSGL 363

QY 372 SIAGITGAKKVTPE-----WASSENQAA--PQ 396

DB 364 -----OKARPE-----QVAKATDTDPADTAK 385

RESULT 15

Q8CVN7 PRELIMINARY; PRT; 385 AA.

AC Q8CVN7: 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Acridavine resistance protein E precursor.

GN ACRE OR C4031.

OS Escherichia coli O6.

DR EMBL; AE011925; AAM37689.1; --
 DR InterPro; IPR006143; H1YD.
 DR Pfam; PF00529; H1YD; 1.
 KM Complete proteome.
 SQ SEQUENCE 404 AA; 42179 MW; F82FD7C04BC7C8B7 CRC64;

Query Match 36.7%; Score 741.5; DB 16; Length 404;
 Best Local Similarity 42.3%; Pred. No. 3.5e-33;
 Matches 174; Conservative 65; Mismatches 151; Indels 21; Gaps 7;

QY 8 AMRAALAAVA--LVLSGCK--GGDAAQGGPAGREAPVAVGVVTHPQVATLTVEL 63
 DB 6 SFRSLVAVAATAVLSACSPPGGPPPEG-----TPKGVLTIVKQOPVLTTEL 57
 QY 64 PGRLESLTADVRAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTYEANLESARAQATATQ 123
 DB 58 PGRTPVYISLRPQVGGIVGSRQFTREGDVKAGGTLQIDPAQVRASTASQAQSLAQE 117
 QY 124 ATLAKADADLARYKPLVAEAVSRQDYDAVTAKRSAEGVKAQAQAIKSAGININRSRI 183
 DB 118 ATLRTAQAKARRYKELAIKAIISOEGDDTDALQAKADVAAGKASVETARINLAFRL 177
 QY 184 TAPISGFGOSKVSSEGLTLLNGDTVLATIRQTNPMYVNVNQSASEVWKLRLQAEGL- 242
 DB 178 DAPISGRIGRSSVTEGALVTANQATALTITQDLDTVIDVTQPSAVLRLOAMARGDLE 237
 QY 243 LAAGVIAVAGIKFDDGTVYPERGRLLFADPVNNESTGQITLRAAVPNDQNTIMPGLYRV 302
 DB 238 QAGGGAANVSLLEBGSVYPIQGHIAFSDVTVDDQTSITLRAVFPNNABLLPMTYRA 297
 QY 303 LMDQVAVDNAAFVPOQAVTR--GAKDTVMIVNAQGMPEPREVTAAQGGTNWIVTSGIKD 360
 DB 298 VLOGVEKQGVILVPOQAVSRNAGKPTAFVVGADHKLQLRLMETREVGDMVRSGLKP 357
 QY 361 GDKVYVNEGISITAGITGAKKVTPEKASSENQAAAPQSGVQTAASEKTASEA 411
 DB 358 GEQVVEGASRARDGIVKTVP--W---QPKATPAAAGNAGPSAPTAPRA 402

RESULT 18

Q8E8H1 PRELIMINARY; PRT; 382 AA.
 AC Q8E8H1;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Multidrug resistance protein, Acra/Acre family.
 GN S04693.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heideberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eisen J.A., Seehadri R., Ward N., Meehe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beanan M., Brinkac J., Daugherty S.,
 Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Imprimis M., Lee K., Berry K., Lee C.,
 Meallier J., Khouri H., Gill J., Utlcherback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AE015901; AAM57652.1; --
 DR TIGR; S04693; --
 KM Complete proteome.
 SQ SEQUENCE 382 AA; 40186 MW; 2BBB85FC4DF09A53 CRC64;

Query Match 36.4%; Score 734; DB 16; Length 382;

Best Local Similarity 44.0%; Pred. No. 8.3e-33;
 Matches 168; Conservative 74; Mismatches 130; Indels 10; Gaps 6;

QY 13 ALAAVALVSSCGKGD-AAQGGPAGREAPVAVGVVTHPQVATLTVELPGLSELR 71
 DB 8 ASVISVALMMMAKSPQEKKAHGG--AGPQSE--VGVIKVEAKQVIKIELPGRSKAFL 63
 QY 72 TADVAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTYEANLESARAQATATAKKDA 131
 DB 64 EAEVRPQVNGIITRSFPEGGNVKGESLYIIDATYKAAVSNADLARANASLASAKA 123
 QY 132 DLARYKPLVAEAVSRQDYDAVTAKRSAEGVKAQAQAIKSAGININRSRIAPISGPI 191
 DB 124 KAARYQVLKTNALISKOPDEADAAVKALASVTAEBAAINTAKINILEYTEVLAPISGRI 183
 QY 192 GQSVSESTLLNAGDTVLATIRQTNPMYVNVNQSASEVWKLRLQAEGLAADGVATV 251
 DB 184 GKSSVTAGALVTANQATALTITQDLDTVIDIAQSSAQLEKAKLQKGLAADNA-DV 242
 QY 252 GIKFDDGTVYPERGRLLFADPVNNESTGQITLRAAVPNDQNTIMPGLYRVIMPOVAVDN 311
 DB 243 QLVIVDGVYHTGKLGFAEVSVDQNTGSVTLRAEFPPDGVLLPMTYRAMLNAGTPQ 302
 QY 312 AFVVPQAVTRGARD--TWMIVNAQGMPEPREVTAAQGGTNWIVTSGIKDGRVVEGI 369
 DB 303 AILVPOKAITENAKGEAVAMVNAQGVKAEKIVTTAEVYNHQMRTVSGLAVGDVIVEGJ 362
 QY 370 SIAGITGAKKVTPEKASSENQ 391
 DB 363 Q-KIRPGAPVPIVVISATQAO 382

RESULT 19

Q9WMZ9 PRELIMINARY; . PRT; 384 AA.
 AC Q9WMZ9;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE TGA.
 GN TTGA.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DOT-T1;
 RX MEDLINE=98317269; PubMed=9642183;
 RA Ramos J.L., Duque E., Godoy P., Segura A.;
 RT "Efflux pumps involved in toluene tolerance in Pseudomonas putida DOT-
 RT T1E.";
 RL J. Bacteriol. 180:3323-3329(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DOT-T1;
 RA Segura A., Duque E., Ramos J.L.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF031417; AAD39553.1; --
 DR InterPro; IPR006143; H1YD.
 DR Pfam; PF00529; H1YD; 1.
 SQ SEQUENCE 384 AA; 41249 MW; 1E1P9BA412D0F6C4 CRC64;

Query Match 36.2%; Score 730.5; DB 2; Length 384;
 Best Local Similarity 40.2%; Pred. No. 1.3e-32;
 Matches 169; Conservative 80; Mismatches 120; Indels 51; Gaps 10;

QY 6 FK-AMRAALAAVALVLSGCKGDAQGGPAGREAPVAVGVVTHPQVATLTVELP 64
 DB 3 FKRAVTRLVSAVALATLISGCK-----EBAAPAAQAPQGVVITIGQAFITSELP 54
 QY 65 GRLESRTADVRAQVGGIIQKRLFOEGSYVRAGOPLYOIDSSTYEANLESARAQATATQA 124

```

Db 55 GRTSAVRVAEVRPQVNGIILKRLFKESGEVKGQQLYQIDPAVYEATLANKANL---A 111
Qy 125 TLAKADADLARYKPLVAAEAVSRQEDAVATKRSAGAEVKAQAQAIKSAGINLRSRIT 184
Db 112 TRSLAE----RYKQIDEGQAVSKQRYDDA-NAKR-----LQAEASLSAQIDLRKYTVL 160
Qy 185 APISGFIFQSKVSEGTLLNAGDTVLATIRQTNPMYVNTQSASEVKKLRQIAEKL-L 243
Db 161 APISGRIRSRSSFTGEGALVNSGQTDAMATIQQLDPYVDVSTQSTALLKRLRDESGQLQK 220
Qy 244 AADGVIAGIKFDDGTVPEKGRLLFPADPVNVESTQITLRAAVPNQNIIMPGLYRVVL 303
Db 221 AGNNAASVQLVLEDSGLFKQEGRELFSEVADETTGSVTLRALFPNPDHTLLPGMFVHAR 280
Qy 304 MDQVAVDPAFVFPQQAATRGAK--DTWMIVNAQGMPEPREVVAQOQGTWMIIVTSGLDKG 361
Db 281 LKAGVNAVAIILAPQGVTRDLKGAFTALVNVQENVEIKRLKASTLGSMDLIEGLNPG 340
Qy 362 DKVVEG-----ISIAGITGAKKVTPEKMASSENOAAAPQSGVQTASEAKTASAE 412
Db 341 DRLITEGQYVRPGVEVKKVSDATNVK-----PAGPDQANAAKADAKAE 384

```

RESULT 20

```

O8PQJ6 PRELIMINARY; PRT; 408 AA.
AC O8PQJ6;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Acridin resistance protein.
GN SMDA OR XAC0327.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camaroto G., Cannavan F., Cardozo J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.V., Cursino-Santos J.R., El-Dorriy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezera R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA Secubal J.C., Kitaajima D.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL; AE011658; AAM35219.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW Complete proteome.
SQ SEQUENCE 408 AA; 42533 MW; 55DD3E271D691265 CRC64;

```

Query Match 36.0%; Score 727.5; DB 16; Length 408;
 Best Local Similarity 41.2%; Pred. No. 2,1e-32;
 Matches 165; Conservative 80; Mismatches 140; Indels 15; Gaps 6;

```

Qy 14 LAAVAIVLSSCGKGDAAQGGPAGREAPAPVGVTVVHQTVALITVELGRLESLFTA 73
Db 18 LPAATVITMAAC-----SSRQPP--QMPQTVGVQVTLKQVRLADVDTLGGRTVAAYTS 68
Qy 74 DVRAOVGGIIQKRLFOEGSVYRAGOPTLYQIDNSTVEANLSEARAQLATAQATLAKADL 133

```

```

Db 69 DVRPQVGIVAKRLFTFGQGVQAOQVLYEIDPASYQAAVDTAKGDLAQAEAAVISAEPKA 128
Qy 134 ARYPLVAABEASQCEVDAVATKRSAGAEVKAQAQAIKSAGINLRSRTAPISGFICQ 193
Db 129 ORYQTLVGLDVAISKQGGDDALATIRSENAVNAKASLQTRIMLDTRITAPISGRIGT 188
Qy 194 SKVSEGTLLNAGDTVLATIRQTNPMYVNTQSASEVKKLRQIAEKLAAADGVIAGVI 253
Db 189 SSTYSGALVSAQSGEVLATINQQLDPYVDVSTQSTALLKRLRDESGQLQK 248
Qy 254 KFDGITYPEKGRLLFPADPVNVESTQITLRAAVPNQNIIMPGLYRVVLMDQVAVDPAF 313
Db 249 QLEDGSTYASGTLLEVDAVDATATGVKRAAVPNPNERLLPQMYVTAKLSMAVDEQAI 308
Qy 314 VVPOQAATRGAKD--TWTMIVNAQGMPEPREVVAQOQGTWMIIVTSGLDKGDKVVEGISI 371
Db 309 LVFPQQAISRSKGAVALVVGSDNKAQORVLTGDAIGDKVVRQGLKAGDKVIVQGLQK 368
Qy 372 AGITGA--KKVTPREMASSENOAA-APQSGVQTASEAKTA 408
Db 369 ASV-GAEVKAVEVTAAMAGNAATAATGTAAPKA 407

```

RESULT 21

```

O8ZC88 PRELIMINARY; PRT; 395 AA.
AC O8ZC88;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Multidrug efflux protein (Acridine efflux pump).
GN ACRA OR YPO3132 OR Y1050.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Pankiewicz M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baeham D., Bentley S.D., Brooks K.C., Cerdano-Tarrega A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jasele K., Karlshav A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Slimmons M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Medisevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Herfsterson J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Berry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AJ141415; AAC92367.1; -.
DR EMBL; AB013708; AAC84631.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW Complete proteome.
SQ SEQUENCE 395 AA; 41889 MW; 8D25E40E22ECBA36 CRC64;

```

Query Match 36.0%; Score 726.5; DB 16; Length 395;
 Best Local Similarity 40.6%; Pred. No. 2.2e-32;
 Matches 164; Conservative 73; Mismatches 142; Indels 23; Gaps 5;

```

Qy 11 AALAAVAIVLSSCGKGDAAQGGPAGREAPAPVGVTVVHQTVALITVELPGRLESL 70

```

```

Db 11 AAILVSSGLVIGC-NDKDAQAQ-----AQAQAEVGVITLKAPLITITDLPKRTSAF 64
Qy 71 RTADVRAVGVGIIQKRLFOEGSVYRAGOPLYOIDSSTEYANLESARAQATLAAD 130
Db 65 RVAEVRPVSGLIKRNYIEGSDVYAGSLYOIDPATYQAADSKGLAKQAQAQIAH 124
Qy 131 ADLARKYPLVAEAVEROEYDAAVTAKRSAAEAGVAKQAQAIKAGININRSRTAPISGF 190
Db 125 LTVNRYKPLLGTNYISKOEYDQALSDAQADATVLAAKAALLESANINLAAYQVRSFISGR 184
Qy 191 IGSQSVSGLTLNAGDTVLATIRQTNPMYVNTOSASEVWKLKROIAEGLKLAADGVI 250
Db 185 TGKSAVTGALVTSQAASAMTVOQLDPMYVDVVTQSSDEFRLKKEKELADGILKQENGKAK 244
Qy 251 VGIKFDDGVYPEKGRLLFADPVNVESTGOITLRAAVPNDQNIIMPGLYVRVLMDOVAVD 310
Db 245 VRLLENGVETETITLESQVTVDETTGSLTLRAIFPNPEALDPGFVRRLDEGIRP 304
Qy 311 NAFVVPQAVTRGAKD--TVMIVNAQGMPEPEVTVAAQOQGTMTVITSLKDKGDKVVEG 368
Db 305 DALVTPQGVTRNPRGEGATAVVAGNRYEMKRTLIAQAIQNKMLVTGGLAKAGRLIISG 364
Qy 369 ISIAGITAKKVTTPKEMASSENQAAAPQSGVOTASEAKTASAE 412
Db 365 L-----QKIKP-----GVEVKVQEVITDIAPETAPADTAK 393

```

RESULT 22

Q8UB14 PRELIMINARY; PRT; 427 AA.

```

AC Q8UB14;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE RND multidrug efflux membrane permease.
GN ACRA OR ATU3203 OR AGR_L_3215.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boye D., St.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-Y., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neeter E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ourlijo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaundin M., Iartchouk O., Epp A., Liu F.,
RA Wolam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AF009250; AAL4019.1; ALT_INIT.
DR EMBL; AB008363; AAK90184.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.

```

KW Complete proteome.
SQ SEQUENCE 427 AA; 45728 MW; 9E7C2E07AC8B9781 CRC64;

Query Match 35.8%; Score 723; DB 16; Length 427;
Best Local Similarity 40.0%; Pred. No. 3.9e-32;
Matches 161; Conservative 83; Mismatches 143; Indels 16; Gaps 7;

```

Qy 11 AALAAVAALVSSCGKGDAAQGGOPAGREAPAP-VGVVTVHPQTVALTVELPGRLES 69
Db 32 AAVLTGLVLY-----GQSDQASAP---APPGAVKVAVAVPELPIINELPGRAP 82
Qy 70 LRTADVRAQVGGIIQKRLFOEGSVYRAGOPLYOIDSSTEYANLESARAQATLAAD 129
Db 83 TRLEAVPRVSGIIVVERFEGSIVKEGDLVLRIDRAPFQVRVDSAEGLTARRAQAQQA 142
Qy 130 DADLARKYPLVAEAVEROEYDAAVTAKRSAAEAGVAKQAQAIKAGININRSRTAPISG 189
Db 143 RQYADROQQLRRSNVSGQGFEDNAILAQAADAEVAEAGVABERMLQYADVAPAPISG 202
Qy 190 FIGQSVSEGLTLNAGDTVLATIRQTNPMYVNTOSASEVWKLKROIAEGLKLAADGVI 249
Db 203 VIGARITEGALVATGSENATIQQLDPIYADTQPADLIRKALQDQQLMTGQNEA 262
Qy 250 AVGIKFDGVYPEKGRLLFADPVNVESTGOITLRAAVPNDQNIIMPGLYVRVLMDOVAV 309
Db 263 EVNLIFDDGSRYYPVSGRLFSEAADVDETTGQVTLRGEPFNPNGDLLPQMYVRVQIQGSIQ 322
Qy 310 DNAFVVPQAVTR--GAKQVMTVNAQGMPEPEVTVAAQOQGTMTVITSLKDKGDKVVE 367
Db 323 KAFAVPQAAVORAGGQASVLYVNAEDVTEQRRVSGRSISGDIWSEGLDDDBRVAAE 382
Qy 368 GISIAGITAKKVTTPKEMASSENQAAAPQSGVOTASEAKTASE 410
Db 383 GFOKTA-PGA-KVPEPMSQEPDVAAA--AGSEGAASESISE 421

```

RESULT 23

Q9KJC3 PRELIMINARY; PRT; 371 AA.

```

AC Q9KJC3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Periplasmic linker protein.
GN APPA
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S12;
RX MEDLINE=21097242; PubMed=11160799;
RA Kleboom J., de Bont J.A.M.;
RT "Identification and molecular characterization of an efflux system
RT involved in Pseudomonas putida S12 multidrug resistance."
RL Microbiology 147:43-51(2001).
DR EMBL; AF183959; AAF73831.1; -.
DR InterPro; IPR006143; HLYD.
DR Pfam; PF00529; HLYD; 1.
SQ SEQUENCE 371 AA; 40276 MW; B86AACB9A6133645 CRC64;

```

Query Match 35.7%; Score 720; DB 2; Length 371;
Best Local Similarity 43.5%; Pred. No. 4.7e-32;
Matches 160; Conservative 74; Mismatches 108; Indels 26; Gaps 8;

```

Qy 6 FK-AMRAAALAAVALVSSCGKGDAAQGGOPAGREAPAPVGVTVHPQTVALTVELP 64
Db 3 FKAVATLVSAVALATLISGCK-----EBAAPAAQAPQGVVITIQOAPFTLSLPL 54
Qy 65 GRLESRTADVRAQVGGIIQKRLFOEGSVYRAGOPLYOIDSSTEYANLESARAQATLAAD 124
Db 55 GRTSAVYRAVAVRPQVNGIILKRLFKESGEVKEGQQLYOIDPAVEATLAAKAKALL--A 111

```

Qy	123	TLAKADADLAKYPLVAAAEVNSRQGYDDAAVTAKRSAGVKAQQAIIKSAQINLNSRRT	184
Db	112	TRSLAE----RYKOLIDEQAVSKOEYDDA-NKR-----LOEASLKSQAQIDRLRTKVL	160
Qy	185	APISGFICQSKSESGTLLNACGDTYVATITROTNPMYVNTQSASEWVKLRROIAEGKL-L	243
Db	161	APISGRITRSSTTEBALTNSGOTDAMATIQQLDPIVVDYTGSTABELTKLRRLDESGQLQK	220
Qy	244	AAQGVYANGIKFDGQTVYPEKGRLLFADBPVNNESGQITLRAAVPNDQIILMGLVYRVL	303
Db	221	AGNNAASVOLVLEDESLFQOEGRLEFSEVAVDETTGVTLRALFPNPDHITLLPGMVFVHAR	280
Qy	304	MDQVAVDAFVFPQQAATRGAK--DTVMVNAQGMPEPREVVAQCOQNTMIITSGLKDG	361
Db	281	LKAGVNAHAAILAPQCGVTRDLKGAPALVYVNDENKELNLQLKASRTLGSMDLIEEGLNG	340
Qy	362	DKVVEGI	369
Db	341	DRLLTEGL	348

RESULT 24
Q92T02
ID Q92T02 PRELIMINARY; PRT; 407 AA

01-DEC-2001 (TReMBLrel. 19, Created)
 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Putative multidrug efflux system protein.
 GN R00193 OR SMC02868.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 NCBI_TaxId=382;
 RN
 RP
 RC
 RX STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Balloy-Hubler F., Gouzy U., Bothe G., Ampe F., Batsu
 RA Roistard P., Becker A., Boultry M., Cadieu E., Dreano S., Gloux
 RA Gorie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Fumeille B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenberg M., Weidner S., Gallibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL: AL591782; CAC14580.1; -
 DR InferPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SW
 SE Complete proteome.
 SC
 QO SEQUENCE 407 AA; 43732 MW; 53FAB3B0FC55042A CRC64;

Query Match	35.6%;	Score 719.5;	DB 16;	Length 407;
Best Local Similarity	40.2%;	Pred. No. 5.7e-32;		
Matches 164;	Conservative 79;	Mismatches 148;	Indels 17;	Gaps 77;

```

QY 9 MRAAALLAAVALV--LSGCGKGDDAAGQGPAREAPAP-vwGVVYVHPTVALTYLTPG 65
Db 7 MRPPLIAAALAVYIFLAGQCKNE-----QAAAAAPPSPPVAVFTTKAPRLPTNLP 61
QY 66 RLESRTADVRAQVGIIQKRLFOEGSSYVACQPLVIDSTYEANLESRAOLATCAQT 125
Db 62 RTAIRIAEVRRIISGIVVERVEQGTWYKESGVLRIDPAPQVAVKSDAELTLKRAQV 121
QY 126 LAKADADLARVYKPLVAAEAVSRQEYDAAVTAKESAEGVKAQAQAIKAGININRSRTA 185
Db 122 VDAQARTADRQSLKEAQVTAQVQYDDALIALAQAADAVGIAAGIAEAKLTNQYTVTA 181
QY 186 PISGFIGQSKVSEBGLTLNAGDTTVLATIKOTNPMYVNTOSASEWMLKRLQIAEGKILAA 245
Db 182 PISGIRGRALLIEGALVNTNDPONTALITQOLDIYVDFOTOSADLLRLRLKALDKQOMMSA 241

```

QY	246	DGVIAVGIKFDGDIYVPEKGRLLFDPVYNSTGQITLRAAPNDONIMLPGLYRVLMD	305
		242	KNKEAEVOLITLDDPYALKGRLLFSEAAVDAITTGVTLLRGEPFNPNDLLPGIYRVVQIQ
Db	306	QYAVNAKVVYVQQAIVR--GAKDTYMIYVAQSGMEPRFETVAQQGOSTNMYLVSGLDGDK	363
QY	302	QGLEKDAITVPOAAVORNNAGSQYVVVNAADNKVFERNYTLGRVGERMQVITSGLRPEK	361
Db	364	VYVEIGISLAGITGAKKVTPEK-----ASSENQAAPASQGVDTASEAK	406
QY	362	VIVSEFQKVG--PGA--PVQPSMDPNAKKPPQASADNGEKPRATEVK	407
Db			

RESULT 25	
Q8GC84	
ID Q8GC84	PRELIMINARY; PRT; 374 AA

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE BeFA lipoprotein precursor.
GN BeFA.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=548;
RX 11)
RN
RP
RC
RC STRAIN=BW16627;
RA Maasi M., Pages J.M., Pradel E.;
RA "identification and characterization of the Enterobacter aerogenes
RT eefABC operon encoding a putative tripartite efflux system.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ508047; CAD48681.1; -.
FT
FT Signal; Lipoprotein.
KW
KW SIGNAL 1
KW CHAIN 18 374 POTENTIAL.
KW CHAIN 18 374 BeFA LIPOPROTEIN.
SQ SEQUENCE 374 AA; 39374 MW; 3332BFD6AFAB942BB CRC64;

Query Match 35.6%; Score 719; DB 2; Length 374;

Match	Conservative	Mismatches	Indels	Gaps
164	65	122	22	7

QY 12 AALAAVALVLSSCGKGGAQGQPAGREAPAPVGVTVHPQTVALTVELPGRLESLR 71.

Db 6 APIATALL-LSGC---DNVQTAP---QQPTPEVGIVTLLQAQPPVVSQLTGRTASL 57

Qy 72 TADVRAQVGCIIQKRLFQEGSVRAGQPLYQIDSSSTEANLESARAQLATIAQATLAKADA 131

DB 58 SAEVRPQVGGI IQKRLIEGDMVKAGQAL IQIDPSSIKAIWNEAAXXALNVAQALVSDCQ II /

132 UAR I AF DV ABE V S N Q E I D P H V I M U U B H B H O V I O R Y G R H T W O R O T I N T I N K O N T I T N I T L I O O L I

Uy

[illegible][illegible]

353 CTYENDCTVBEKCP.I.EADRYNESTCOTI.PAAYPNONTIMBG.YVYVIMDOVAVDN 311

```
Db      234 TLLEDGTYPEKGRALTTEVAVDSTGSVTLRIFPNPOHVLPGMFVRARIDEGIMND 293
```

312 AFVVPQAVTRGAKD--TWMI VNAQGGMEPREVTVAQQQGTNMI VTSGLKDGDKVVEGI 369

Db 294 AILAPQGITRDAGDATAALVVDADNNVEQRTVQTGDTYGDKWLVLGLKAGDKLIVEG- 352

QY 370 SIAGITGAKKVT 382

D6 353 -----TG--KVSP 358

RESULT 26

08CM42
ID 08CM42 PRELIMINARY; PRT; 384 AA.
AC 08CM42;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Acriflavine resistance protein A precursor.
GN C1761.
OS *Escherichia coli* O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.,
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic *Escherichia coli*."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016760; AANB0227.1; -.
KW Complete proteome.
SQ SEQUENCE 384 AA; 41129 MW; FC9E69E4583EC7B8 CRC64;
Query Match 34.7%; Score 701.5; DB 16; Length 384;
Best Local Similarity 42.5%; Pred. No. 5.1e-31;
Matches 162; Conservative 69; Mismatches 133; Indels 17; Gaps 7;
QY 7 KAMRAALAAVALVLSGCKGDAAGGQAGREAPVAVVTVHPQTVALTVLPGLES 66
DB 10 KIMKVIATSVYAMLLSGC-----DNTQSNNSPSETE---VGVTTLSQPVSVSELTGR 62
QY 67 LESRTADVRAQVGGIIOKRLFOGGSYVRAGOPLYOIDSSTYENLSARAQATLAQATL 126
DB 63 TSAALSLEVPQVGIIQKRLFEKGDVAKAQPLYQIDAAASYQAAMNEARALQAOA-L 121
QY 127 AKAAADLA-RYKPLVAEAVSROEYDAVTAKRSAEGVKAQAQAIKSAGINLRSRITA 185
DB 122 VKACQKQKQRTYLVKENGVSQDADAGSTCAODKASVEKKKALETARINLDMVTYTA 181
QY 186 PISGFIQSKVSEGLTLNAGDTVLAITRTQNPVYVNTQASAEVWKLRQIAEGKLLAA 245
DB 182 PISGRIGISSTVPALVYASODTALTIRGLDMYVDITRSSVDLRLKCO----SLATN 237
QY 246 DGVAVGIRKEDGTVPYKGRLLPADPVNVESTQITLRAAVPNDONILMGLYVRVIMD 305
DB 238 SDTWSVSLIEDGTTYSKGRLELVEAVDESTGVTLRATFPPRQQLLPGMFVRARVD 297
QY 306 QVAVDANFVVPQAVTRGARD--TWMIYVNAQGMPEPEVYVAAQOGTNWYVTSGLKDGK 363
DB 298 EGWVEDAILPQOGVTRDAKNALVYVNDKNVEORTLETGETYGRKWLVLNGLHNGDR 357
QY 364 VVEGISIAGITGAKKVTPE 384
DB 358 LIVEG--SAKYTSQYKAVE 376

RESULT 27
Q8RSM2
ID 08RSM2 PRELIMINARY; PRT; 388 AA.
AC 08RSM2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE MexC protease.
GN MEXC.
OS uncultured bacterium.
OC plasmid pB4.
OC Bacteria; environmental samples.
OX NCBI_TaxID=71133;

RN [1]
RP SEQUENCE FROM N.A.
RA Tauch A.;
RT "The Incp beta plasmid pB4 encodes a tripartite antibiotic efflux
system of the RND-MFP-OEP type conferring erythromycin and
tetracycline resistance in *Pseudomonas* sp. B13."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ431260; CAD24338.1; -.
DR InterPro; IPR06143; HLYD.
DR Pfam; PF00529; HLYD; 1.
KW plasmid.
SQ SEQUENCE 388 AA; 41104 MW; 83CE3D1E04BF737 CRC64;
Query Match 34.7%; Score 700.5; DB 2; Length 388;
Best Local Similarity 40.6%; Pred. No. 5.9e-31;
Matches 159; Conservative 81; Mismatches 131; Indels 21; Gaps 8;
QY 12 AALAAVALVLSGCKGDAAGGQAGR--EAPVAVVTVHPQTVALTVLPGLES 69
DB 11 SVISLVITMTLAGC-----DQPAEWGEAPPREVDAVKTPEPTVIALPGR1EP 61
QY 70 LRTADVRAQVGGIIOKRLFOGGSYVRAGOPLYOIDSSTYENLSARAQATLAQATLAKA 129
DB 62 VRAVEARAVAGIVLSRKFEGADVKAGDVLFLIDPAPFKAALSRAEQLAQADAOLEFOT 121
QY 130 DADLARYKPLVAEAVSROEYDAVTAKRSAEAGVKAQAQAIKSAGINLRSRITAPISG 189
DB 122 QATYHRYEPVLYKIDAVRODPVATALQSKAKKRAQADLETARLDLGATYRAT1AG 181
QY 190 FIGQKVSSEGLTLNAGDTVLAITRTQNPVYVNTQASAEVWKLRQIAEGKLLADG-V 248
DB 182 RIGRAQTEBALVQGGESTLLARIQUDPVADFTQPAADALRLRAIAEGKVGDDGKA 241
QY 249 IANGIKRDKDQTVPEKRLFPADPVNVESTQITLRAAVPNDONILMGLYVRVIMQVA 308
DB 242 LLSLV--DQTSLSRGLTFTVYVDRSTQIALRGRFDPNPGVLLPQWVVRTRTQGN 298
QY 309 VDNAFVVPQAVTRG--AKDTWMIYVNAQGMPEPEVYVAAQOGTNWYVTSGLKDGKVV 366
DB 299 SD-AIIVPQAVVRAIDGQASVWVLGAEIVTVEARPYTTGVMQGRWQITBGLKGDVIV 357
QY 367 EGISAGITGAKKVTPE-EWASSENQAAPOS 397
DB 358 S--SLAIIQGVTVVPRQGAQAPAAASQPA 387

RESULT 28
Q51395
ID 051395 PRELIMINARY; PRT; 387 AA.
AC 051395;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Membrane fusion protein MEXC (RND multidrug efflux membrane fusion
protein MEXC).
GN MEXC OR PA4599.
OS *Pseudomonas aeruginosa*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; *Pseudomonadales*.
OC Pseudomonadaceae; *Pseudomonas*.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97032139; PubMed=8878035; Zhao Q., Wada A., Yamasaki T.,
RA Poole K., Gotch N., Tsujimoto H., Li X.Z., Nishino T.,
RA Nishit S., Yamagishi J., Li X.Z., Nishino T.,
RT "Overexpression of the mexC-mexD-opr efflux operon in nfxB-type
multidrug-resistant strains of *Pseudomonas aeruginosa*."
RL Mol. Microbiol. 21:713-724(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Watterner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 DR EMBL: U57969; AAB41956.1; -
 DR EMBL: AE004873; AAC07987.1; -
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR Complete proteome.
 KW SEQUENCE 387 AA; 40838 MW; 8CF5BF3B2314A0A CRC64;

Query/Match 34.6%; Score 699.5; DB 16; Length 387;
 Best Local Similarity 41.5%; Pred. No. 6,6e-31;
 Matches 162; Conservative 82; Mismatches 131; Indels 15; Gaps 9;

QY 10 RAALAAVAALVLSGCKGDAAGGQAPAGREAPAPVGVTVHPQVATLTVLPGRLES 69
 Db 9 RIGALAAIA-LACG-PAERQEA-EMVLP-VEVLTVQAEPLALSELPGRIEP 60
 QY 70 LRTADVRAOVGIIQRLFOEGSYVRAGQPLYOIDSSTYEANLESARAQLATQAQTLAKA 129
 Db 61 VRVAEVRARVAGIVVRKFEAGADYKAGDLFQIDPAPLKAASRABELANRAVLREA 120
 QY 130 DADLARIYPLVAEVSROEYDAVTAKRSAAGVKAQAQAIKSGINLNSRTIAPISG 189
 Db 121 QARVRYEPLVKIQAVSOODPDTATADRSAAEARSQAQDLERLNLGVAASVTAPISG 180
 QY 190 FIGOSKSEGLTLNAGDTTVLATIRQTNPMVYNVTQASSEWKLRLQIAEGLLAADVI 249
 Db 181 RIGRLVTEGALVGGSEATLMARIQQLDPTIADFTQIAEKLRLDAIKGTLAAGDSQ- 239
 QY 250 AVGIFDDGTYVPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYVRVLMDOYAV 309
 Db 240 ALTLRV-EGTYERGALQFADVAVDRTGQIALKGFANPDGVLTPGMVYRVRTPGID 298
 QY 310 DNAPFVPOOAVTRGAKDT--VMIVAAQGMPEPRTVAQOQGTNMTVTSGLKDGKVV 367
 Db 299 NQAILVPGRAVRRSSDGAQVWVAGADRAARSRSVGTSGMOSRWQITGEGLEPGDRIIVG 358
 QY 368 GISISAGITGAKKVTPE-EMASSENQAAPQ 396
 Db 359 G-LAAVOPGVKIVPKPGQAQAQASPAPO 386

RESULT 29

OS87E1 PRELIMINARY; PRT; 373 AA.
 AC OS87E1
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative efflux pump (Putative multidrug-efflux transport
 DE protein).
 GN Z2509 OR ECS1863.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDU933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RT Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL: AE005377; AAC56521.1; -
 DR EMBL: AP002556; BAB35286.1; -
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR Complete proteome.
 KW SEQUENCE 373 AA; 39612 MW; 2A573537AB44CDBF CRC64;

Query/Match 34.5%; Score 697.5; DB 16; Length 373;
 Best Local Similarity 42.7%; Pred. No. 8.1e-31;
 Matches 162; Conservative 66; Mismatches 134; Indels 17; Gaps 7;

QY 9 MRAALAAVAALVLSGCKGDAAGGQAPAGREAPAPVGVTVHPQVATLTVLPGRLE 68
 Db 1 MKYIATSVAMLLISG-DMTQSNSSPSEIE--VGVVTKSQPVSVSLTRTS 53
 QY 69 SLRTADVRAOVGIIQRLFOEGSYVRAGQPLYOIDSSTYEANLESARAQLATQAQTLAK 128
 Db 54 AASAEVRPVGIIQRLFKREGDLVKAQPLYOIDAASVOAMNENRALLQAQQA-LVK 112
 QY 129 ADADLA-RYRPLVAEVSROEYDAVTAKRSAAGVKAQAQAIKSGINLNSRTIAP 187
 Db 113 ADCKAKRYRPLVAVKENGVSQDDADDAQSTCAQDASVAAKKAALERLNLDMWTIAP 172
 QY 188 SGFIGOSKSEGLTLNAGDTTVLATIRQTNPMVYNVTQASSEWKLRLQIAEGLLAADG 247
 Db 173 SGRISSVTPGALVTVASQDTALTTRIGDPTVVDLTRSSVDLLRLRKQ----SLATNSD 228
 QY 248 VIAVGIFDDGTYVPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYVRVLMDOY 307
 Db 229 TMSVSLILEGGTYVSEKGRLELFEVAVDESGTVTLAIPNPPOQLPMPFARARVDEG 288
 QY 308 AVDAFVPOOAVTRGAKD--TWIVAAQGMPEPRTVAQOQGTNMTVTSGLKDGKVV 365
 Db 289 VMDAILIAPDOOQVTRDKAGNATLAVKXNKVREORTLETGETYGDKVLVINGLHSGDRLI 348
 QY 366 VEGISAGITGAKKVTPE 384
 Db 349 VEG--SAKVTSGQTVKAVE 365

RESULT 30

OS93K41 PRELIMINARY; PRT; 398 AA.
 AC OS93K41
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Acra protein.
 GN ACRA.
 OS *Klebsiella pneumoniae*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Klebsiella*.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K23;
 RA Domenech-Sanchez A., Alberti S., Martinez-Martinez L., Garcia I.,
 RA Benedi V.J.;
 RT "Characterization and role in antimicrobial resistance of the AcraB
 efflux pump of *Klebsiella pneumoniae*.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ318073; CAC41008.1; --
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR SEQUENCE 398 AA; 42426 MW; 57CFA07BE450669 CRC64;

Query Match 33.8%; Score 683; DB 2; Length 398;
 Best Local Similarity 39.3%; Pred. No. 5,6e-30;
 Matches 155; Conservative 80; Mismatches 133; Indels 26; Gaps 7;

QY 11 AAALAAALVALVSSCGKGGDAAGGQPRGAPVAVGVVYVHPOTVALTYELPERLESL 70
 DB 11 AVLTMLSSSLALVTC-DKPAQQAQ-----HMEVGIYVTKAVALDITTELPERSAY 63
 QY 71 RTADVRAQVGGIIQKRLFOEGSVYRAGQPLVQIDSTYEANLESAPQATATQATLAKAD 130
 DB 64 RIAEVRPQVSGIILKKNFVSGSDIQAGVFLVQIDPATYQSGYDAKGLAQQAAMNDQ 123
 QY 131 ADLARKYPLVAEAVSRQDYDAVYAKRSAGVKAQAQAIKSGININRSRTAPISGF 190
 DB 124 LTVKRYOKLTQYISQDDYPAVATAQNNAAVAAATVAVETARINLAATYQ-KSPLRSA 182
 QY 191 IGQSK--VSEGTILNAGDTVLATITROTINPMYVNVVQTSASFVMTLRQIAGSKILLADGV 248
 DB 183 AGSGSPVTEBALVQNGQTLATVQVDPVYVDTQSSNDPLRKLQELADARLKQENGK 242
 QY 249 IAVGIKFDGTVYPEKGRLLFADPVNVNESTGQITLRAVPNDONILMPGLVYRVLMDOVA 308
 DB 243 AKVELVYNDGAKYQSGTLFSDVTVDTQTSITLRAFPNPDITLPGMVRARIEGCI 302
 QY 309 VDNAPVFPQAVTGAKD--TWMTVNAQGMPEPREVTAAQOQGNWIVTSGIKDQDKVV 366
 DB 303 NPDALVAVQGVTRTPRGDASVMVVGEGDKVEVRQVTSQAQIGQKMLVTDLKSGDRYIV 362
 QY 367 EGISTAGITGAKKVT-----KEMASSENOAAA 394
 DB 363 -----TGQIKIKPGVQVAKAEVASDDKQAAA 388

RESULT 31

Q9PBP7 PRELIMINARY; PRT; 408 AA.
 AC Q9PBP7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Precursor of drug resistance protein.
 GN Xf2093.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Agencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Faciniani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.U., Hohnsbeil J.D., Junqueira M.L., Kemper E.L., Kltajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos B.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA da Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.",
 RL Nature 406:151-159(2000).
 DR EMBL: AE004025; AAF84892.1; --
 DR InterPro: IPR006143; HLYD.
 DR Pfam: PF00529; HLYD; 1.
 DR Complete proteome.
 SO SEQUENCE 408 AA; 43636 MW; E08A62B9F3BD8EAB CRC64;

Query Match 33.2%; Score 670; DB 16; Length 408;
 Best Local Similarity 36.8%; Pred. No. 3e-29;
 Matches 146; Conservative 90; Mismatches 149; Indels 12; Gaps 5;

QY 9 MRAAALAAVALVSSCGKGGDAAGGQPRGAPVAVGVVYVHPOTVALTYELPERLE 68
 DB 17 LRVIGLAVYVLLALVATCSK-----SEQP---QMPPEVSIYAKPEQTPPIQLDLVGRLS 67
 QY 69 SLRTADVRAQVGGIIQKRLFOEGSVYRAGQPLVQIDSTYEANLESAPQATATQATLAK 128
 DB 68 AYSADVRAVAGIILQKRIITTEGTEVKEGQPLVQIDPATYQSGYDAKGLAQQAAMNDQ 127
 QY 129 ADLARKYPLVAEAVSRQDYDAVYAKRSAGVKAQAQAIKSGININRSRTAPIS 188
 DB 128 AKIVADBARLSPQYIYSRSDVONAEAEARTAAASYEGQAAMQNNRINLVANVTAPIS 187
 QY 189 GFTGQSVSEGTILNAGDTVLATITROTINPMYVNVVQTSASFVMTLRQIAGSKILLADGV 247
 DB 188 GLAQGQGVTEGALVSGSSSTLLTVDQIDPLVNFMSNDELQOLRQAOHQSGVQSSDN 247
 QY 248 VIAVGIKFDGTVYPEKGRLLFADPVNVNESTGQITLRAVPNDONILMPGLVYRVLMDOV 307
 DB 248 TSTITDVLSDGSKYFPGIIGDFSGATVDPSTGVSLRAVLPNDHLLPAPVTFKXNLG 307
 QY 308 AVDNAPVFPQAVTGAKD--TWMTVNAQGMPEPREVTAAQOQGNWIVTSGIKDQDKVV 366
 DB 308 QRNNTTILPGAVQGRDANSVAVVIDKGVKIKNTTIGMKONMILITGKSGDRYIV 367
 QY 367 EGI-STAGITGAKKVTPEKEMASSENOAAAPQSGVQTA 402
 DB 368 DGLQKVEGAPAKAIPKPDLISSSAHAAAKTPHTTA 404

RESULT 32

Q9RBY9 PRELIMINARY; PRT; 398 AA.
 AC Q9RBY9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE SmeA.
 GN SmeA.
 OS Xanthomonas maltophilia (Pseudomonas maltophilia) (Stenotrophomonas
 OS maltophilia).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Stenotrophomonas.
 OX NCBI_TaxID=40324;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UJA-511;
 RA Li X.-Z., Zhang L., Poole K.;
 RT "A multidrug efflux system, SmeRSABC, of Stenotrophomonas
 maltophilia.",
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF173226; AAD51344.1; --
 DR InterPro: IPR006143; HLYD.

DR Pfam; PF00529; HlyD; 1.
SQ SEQUENCE 398 AA; 41888 MW; 2D676251E2C00A9B CRC64;

Query Match 32.8%; Score 662; DB 2; Length 398;
Best Local Similarity 38.0%; Pred. No. 7.9e-29;

Matches 147; Conservative 85; Mismatches 135; Indels 20; Gaps 6;

```

OY 14 LAALVALVSSCGKGDAAAGGQAPGR-EPAP-PVVGVTVPQTALVTELPGRLSESLR 71
DB 14 LLLPLLLALACS-----ADRTDAPAMPEVGVIIASAQPLALQOITLPGAAVFE 62
OY 72 TADVAQVGGIIQKLFQEGSYVRAQGLYQIDSTYEANIESARAQIATQAATLAKADA 131
DB 63 ISEVAPQIGLIRQLFTFEGQGVKAGQLLYQVDPAPYQAPFTARGQLAQAATVLSAP 122
OY 132 DLARKPELVAAEAVROEYDAVATKRSAGEGVKAAQAIKSAGINLRSTRITAPISFI 191
DB 123 KAERTFRAVSMAPKADADDTSAIKQAQANVIAARAALQARINLDTYRTVTAIDRI 182
OY 192 GQSKVSEGTLLNAGDTVLTATIRQTNPMVYVNTQSASEVMKLRQIAEGKLLAADGVIAV 251
DB 183 GTSSVTAGALVAAAGDDTLTTIGRLDPYLDVTSSTQMLARKLDAAGLVKADGKAQV 242
OY 252 GIKPDDGVVPEKGLLPADPVNVESTGQITLRAAVPNDQNLMEGLYVRILMDQAVDN 311
DB 243 KVLLEDGSTVYHEGTLFEVGSADVDPGTNVKLRVAVIPNDGILLPGMYLKVLPMATDAR 302
OY 312 AFVPOQAVTGG--AKDTVMIVNAOGMEPREVTVAQOQGTNMTVTSGLKDGDKVYV-EG 368
DB 303 ALLVPOKAVNREGERPELRLIDAKDHVERRVSTGVVAGQWQITTSGLKAGERVIVSNG 362
OY 369 ISAGITGAKKVPKEMASSENOAAP 395
DB 363 SAVSLGQGVKAVP-----TTAQLAAMP 385

```

RESULT 33

O8X4L0 PRELIMINARY; PRT; 385 AA.

```

ID O8X4L0 PRELIMINARY; PRT; 385 AA.
AC O8X4L0:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN YHU OR 24926 OR EGS4393.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Poser G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocitis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.,
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
RL Nature 409:529-533(2001).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12,"
RL DNA Res. 8:11-22(2001).
DR EMBL; AF005576; AAC58654.1; -

```

DR EMBL; AP002565; BAB7816.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KW Complete proteome.
SQ SEQUENCE 385 AA; 41220 MW; 6823EF67D6141102 CRC64;

Query Match 32.4%; Score 655; DB 16; Length 385;
Best Local Similarity 38.5%; Pred. No. 1.8e-28;

Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

```

OY 25 CGKGDAAAGGQAPGRAPVVGVTVPQTALVTELPGRLSESLRTADVAQVGGIIQ 84
DB 14 CGAMITACDDKSAENTAMTPEVGVTLSPGSVNVVSELPERITVPEVAETRPQVGGII 73
OY 85 KRLFOEGSYVRAQGLYQIDSTYEANIESARAQIATQAATLAKADADLARYKELVAAEA 144
DB 74 KRNFIKEDKVNQGSLLQIDPAPLQALNBSKGLAALSTASNAKRTTFNRQASLLKTNV 133
OY 145 VSRQEDYAAVTAKRSABGVKAAQAIAKSAGININRSRTAPISGFTQSKVSEGTLLNA 204
DB 134 VSRQDYPTARTQLNEAEANVTAKAAVEQATINLQVANVTSPINGVSKSVTGALVTA 193
OY 205 GDTVLTATIRQTNPMVYVNTQSASEVMKLRQIAEGKLLAADGVIAVIGIKDDGVVPEK 264
DB 194 NQADSLVTVOGLDPIYDLTQSVODFLRKEEVASGQIKOVQSTPVQNLNENKRYST 253
OY 265 GRLFPADPVNVESTGQITLRAAVPNDQNLMPGLYVRILMDQAVDNAFVVPQAVTGA 324
DB 254 GTLFSFDPYDETTGSLTIRLIFPNPGDILPGMYTALVDEGRQVAVLVPOGCVTHNA 313
OY 325 --KDTVMIVNAOGMEPREVTVAQOQGTNMTVTSGLKDGKRVVVGIS--TAGITGAKV 380
DB 314 QKATFALILDKDDVVKRLREIASKAIGDQWVVTSGLAQDPRVIVSGLRIRPGIKARIS 373
OY 381 TPKEWASSENO 391
DB 374 SSOENASTESK 384

```

RESULT 34

O8CVL1 PRELIMINARY; PRT; 385 AA.

```

ID O8CVL1 PRELIMINARY; PRT; 385 AA.
AC O8CVL1:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical lipoprotein yhuP precursor.
GN YHU OR C4324.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rosach P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Moley H.L.T., Donnenberg M.S., Blattner F.R.,
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AF016768; AAN82760.1; -
KW Hypothetical protein; Lipoprotein; Complete proteome.
SQ SEQUENCE 385 AA; 41160 MW; 3DC65B6CCF51CBB8 CRC64;

```

Query Match 32.4%; Score 655; DB 16; Length 385;
Best Local Similarity 38.5%; Pred. No. 1.8e-28;
Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

OY 25 CGKGDAAAGGQAPGRAPVVGVTVPQTALVTELPGRLSESLRTADVAQVGGIIQ 84


```

Db 14 CGAMLTACDYSAENAAAMPPEVGVTVLSPGSVNVVLSLPGRTVPYVEAIRPOVGGIII 73
Qy 85 KRLFOEGSVYRAGOPLYOIDSSTYEANLESARAOIATQAATLAKADLAEKPLVAEA 144
Db 74 KRNFBEGKVNQSGSLYOIDPAPLOAEINSAKSLAKSLASTASNRITFPNQASLTKINY 133
Qy 145 VSRQETDAVTAKRSAEAGVKAQAQAIKASAGINLRSRTIAPISGFIGOSKVSSEGLINA 204
Db 134 VSRQDYDPAFTQVLEAEANVTVAKAABVQATINLOYANVTSPITGVSGSKSVTGAIVTA 193
Qy 205 GDTVLVLTIRGTNPMYVNVVQOSASEVMKLRQIAEGKLLADGVIAVIRKDDGTVPPEK 264
Db 194 NQADSLVTVQRLDPIYDVLTSVODFLMKKEVASGQIKQVQSGTPVOLNLENKRGQT 253
Qy 265 GRLLFADPVPVNESTGQITLRAAVPNDQNIIMPGLYVRVIAVDQVAVDPAFVVPQOAVTRGA 324
Db 254 GTLFKSPDPTVDETTGSVTLRAIFPNPNDDLLPGMVTVALVDEGRQVNLVLPQSGVTHNA 313
Qy 325 --KDTWMTVAAGMEPREVTVAOQGTNMTVITSGLKDGKGVVVGIS--TAGITGAKV 380
Db 314 QGKATALLDQKDVQVLEIEASKAIGDQVVTSGLOQGDHVIVSGLQRIKRAAIS 373
Qy 381 TPKEWASSENQ 391
Db 374 SSQENASTESK 384

RESULT 35
ID 09KRV5 PRELIMINARY; PRT; 382 AA.
AC 09KRV5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Peptidase linker protein.
GN TIGD.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2011535; PubMed=10648517;
RA Mosqueda G., Ramos J.L.;
RT "A set of genes encoding a second toluene efflux system in Pseudomonas
RT putida DOT-T1 is linked to the tod genes for toluene metabolism.";
RL J. Bacteriol. 182:937-943(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Mosqueda-cano G.;
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL, Y19106; CAB72258.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
SQ SEQUENCE 382 AA; 41584 MW; EE2C751979592757 CRC64;

Query Match 32.4%; Score 654; DB 26; Length 382;
Best Local Similarity 39.0%; Pred. No. 2,1e-26;
Matches 158; Conservative 70; Mismatches 143; Indels 34; Gaps 10;

```

```

Db 164 PITGRIRGSATVTEGALVTNQAQELATVTOIDPIYVNVNQPITRLGKRALESGRLORV 223
Qy 245 ADVIAVIGFEDGTVPPEKGRLLFADPVPVNESTGQITLAAVNDQNIIMPGLYVRVIA 304
Db 224 GDNQAVSLTLDOSTPYPLGVLFKFSVSAVSTGSVTLAEFPNPHKLLPGFVALL 283
Qy 305 DOVAVDPAFVVPQOAVTRGAK--DTWMTVAAGMEPREVTVAOQGTNMTVITSGLKDG 362
Db 284 NEGEQQAAILVPHQAVGRDARGVPTVAVPDMVRESREVOQTQVGNAMLLAGINDGE 343
Qy 363 KVVVEGISIA--GITGAKVTPKEWASSENQAAAPQSGVQTASEA 405
Db 344 RVITEGVQLARSGI----KVP--VAKKVKMLSEFSGSVQAQA 381

RESULT 36
ID 08P876 PRELIMINARY; PRT; 436 AA.
AC 08P876;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Multidrug resistance protein.
GN ACRA OR XC2367.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camnava F., Cardozo J., Chamberggo F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A.M., Sena J.A.D., Silva C. de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Secudal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL, AE012345; AAA41645.1; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KM Complete proteome.
SQ SEQUENCE 436 AA; 44879 MW; EB45096A5C9CE011 CRC64;

Query Match 32.1%; Score 647.5; DB 16; Length 436;
Best Local Similarity 35.0%; Pred. No. 5,6e-28;
Matches 150; Conservative 83; Mismatches 154; Indels 41; Gaps 6;

```

QY 188 SGFIGQSVSEGTLLNAGDTTLATIRQTNPMYVNTQSASEVWKLRLQIAEGKL-LAAD 246
 Db 176 TGRAGIQVTEGALVSGSEATLLTVDQIDPLVYFAMSSSELAALRQAGSGNVQISGD 235
 QY 247 GVIAVGIRFEDGTVPKGRLLFPADVNESTGQITLRAAPNDONILMPGLYRVLMDQ 306
 Db 236 GKSTTNVELNGTQYHPHPTGLDVSATVDPSTGAVSLRATLPNPMALLPGAFTVFKASL 295
 QY 307 VAVDNAFVFPQAAVTRGAKDT-VMIVNAQGMPEPREVTVAQOQGNMTVTSGLKDGDKV 365
 Db 296 GORNNAYIVPQAAQORDATGAVLALVGDGKVRKGNLTVDQOQKQWITVHMAFGDQVI 355
 QY 366 VEGISA-----GTTGAKVTPKEMASSENQ-----AAPQ 396
 Db 356 VDGVOKAKEGQPAKGPWDPNKPAGQGFAPGAFAAGAGTGGAGAPAGAKADAAAAP 415
 QY 397 SGVOTASE 404
 Db 416 AGEOPKSD 423

RESULT 37

Q93E20 PRELIMINARY; PRT; 396 AA.
 AC Q93E20;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE AdeA membrane fusion protein.
 GN ADEA.
 OS Acinetobacter baumannii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 OX NCBI_TaxID=470;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BM4454;
 RA Magnet S., Courvalin P., Lambert T.;
 RT "Characterization of a RND type efflux pump involved in aminoglycoside
 RT resistance in Acinetobacter baumannii clinical isolate."
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF370885; AAL14439.1; --
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 396 AA; 43376 MW; CABFCD853472F80B CRC64;

Query Match 32.0%; Score 647; DB 2; Length 396;
 Best Local Similarity 38.3%; Pred. No. 5.3e-28;
 Matches 152; Conservative 75; Mismatches 156; Indels 14; Gaps 5;

QY 14 LAAAVALLVSSCGKGDAAGGQPAAGREAPAPVGVTVHPQTVALTVLPGRLSRLTA 73
 Db 9 LFLSLGILLQGC-----DSKEVQAEPPEPAKVSLLSIQPOSVNFSENLPKRVIAFRTA 61
 QY 74 DVRAQVGGIIQKRLFOEGSYVRAQGPLYQIDSTYEANLESARQALATAQATLAKADL 133
 Db 62 EIRPQVGGIIEKVLFPKQSEVRAQALYKINSETFEADVNSVRASLNAEAEVARTKQVL 121
 QY 134 ARYKPLVAAEAVSRQEDAAVTAKSAEAGVAAQAAIKSAGININRSRTIPISGFIGQ 193
 Db 122 ERYEQLPSNMIISKQVSNAAQAYRQALADVAQMAALLARONLNIQYATVRAPISGRIGQ 181
 QY 194 SKVSEGTLLNAGDTTLATIRQTNPMYVNTQSASEVWKLRLQIAEGKL-LAADGVIAVG 253
 Db 182 SFVTGALVGGDTTMTATIQIDKYVDVKQSVSEYERLQALOSGE-LSANSKIVRI 240
 QY 254 KFDGTVVPEKGRLLFADPVNESTGQITLRAAVENDONILMPGLYRVLMDQVADNAF 313
 Db 241 TNSHQOPYNVAKMLFEDINVDPEGTGVTFRLEVNNTERKLLPGMYAVNIDRASIPQAL 300
 QY 314 VVPQAAVTR--GAKTVMIVNAQGMPEPREVTVAQOQGNMTVTSGLKDGDKRVVVGSI 371
 Db 301 LVPQAQAIORNISGEPOVVVINAQGTAEIRPIEIGQYEOFPYANKGLKVGGRVVEGIE- 359

QY 372 AGITGAKVTPKEMASSENQAAAPSGVOTASEAK 406
 Db 360 -RIKPNCKLLAAWKAPAVANHASVETKTSIAEGAQ 395

RESULT 38

O31099 PRELIMINARY; PRT; 382 AA.
 ID O31099;
 AC O31099;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Periplasmic linker protein.
 GN SRPA.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S12;
 RX MEDLINE=98079031; PubMed=9417051;
 RA Kleboom U., Dennis J.U., de Bont J.A., Zylstra G.J.;
 RT "Identification and molecular characterization of an efflux pump
 RT involved in Pseudomonas putida S12 solvent tolerance."
 RL J. Biol. Chem. 273:85-91(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S12;
 RX MEDLINE=99069360; PubMed=9852029;
 RA Kleboom U., Dennis J.U., Zylstra G.J., de Bont J.A.;
 RT "Active efflux of organic solvents by Pseudomonas putida S12 is
 RT induced by solvents."
 RL J. Bacteriol. 180:6769-6772(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S12;
 RA Kleboom U., Dennis J.U., Zylstra G.J., de Bont J.A.M.;
 RL Submitted (Feb-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF029405; AAD12178.1; --
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 382 AA; 41429 MW; 8B7AB0A2DC4C440A CRC64;

Query Match 31.7%; Score 640; DB 2; Length 382;
 Best Local Similarity 36.0%; Pred. No. 1.2e-27;
 Matches 143; Conservative 84; Mismatches 138; Indels 32; Gaps 8;

QY 7 KAMRAAALAAVALLVSSCGKGDAAGGQPAAGREAPAPVGVTVHPQTVALTVLPGRL 66
 Db 8 RALAVITPLTA--LMLISGCGE-----KEQVSATPPPVGVTVRAQALLTITDLPGR 58
 QY 67 LESLRTADVRAQVGGIIQKRLFOEGSYVRAQGPLYQIDSTYEANLESARQALATAQATL 126
 Db 59 TSARFVVEVPPQVSGIIQKRSFVEGAEVKLGQQLYQIDPRYEBQLRRAEANRRTSAQ-NL 117
 QY 127 AKAPADLARYKPLVAAEAVSRQEDAAVTAKSAEAGVAAQAAIKSAGININRSRTIP 186
 Db 118 AR-----RIETLTKTAVSKQQYDDDLAAWKAQAEAYQVAR-----IDVQTRVLSP 164
 QY 187 ISGFIGQSVSEGTLLNAGDTTLATIRQTNPMYVNTQSASEVWKLRLQIAEGKL-LAA 245
 Db 165 ISGRIGSTVTEGALVINGAQSLATVQIDPRIVDVDTOPITKLLGLQKALSGRLQXTG 224
 QY 246 DGVIAVGIRFEDGTVPKGRLLFADPVNESTGQITLRAAVENDONILMPGLYRVLMD 305
 Db 225 ENQAEVSLITDDGSAVPLPGTLKFSVSDPTGQSVTLRAFPNPKLLPGMFAHLK 284
 QY 306 QVAVDNAFVVPQAAVTRGAK--DTVMIVNAQGMPEPREVTVAQOQGNMTVTSGLKDGDK 363
 Db 285 EGVNAAILVPPQAQISRDTRGVPSVWVVKADNTVESREIOTLRTKVGAMALISNGVTGER 344

Qy 364 VVEGIS-----IAGITGAKKVTPEKMASSSENQAAA 394
 Db 345 IITEGVORVRSIGIAVNAVEAKVNLVDPGFAATTEASA 381

RESULT 39

Q8PJN2 PRELIMINARY; PRT; 435 AA.
 ID Q8PJN2
 AC Q8PJN2;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Multidrug resistance protein.
 GN ACRA OR XAC2498.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 CX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 RA Canarotte G., Camarvan F., Cardozo J., Chamberg F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Guber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Mota L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
 RA Trindade dos Santos M., Trufi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima M.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RT "most specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AE011888; AA037349.1; -;
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KW Complete proteome.
 SQ SEQUENCE 435 AA; 4959 MW; CEB07451276FEB6D CRC64;

Query Match 31.5%; Score 636; DB 16; Length 435;
 Best Local Similarity 36.1%; Pred. No. 2.4e-27;
 Matches 151; Conservative 79; Mismatches 162; Indels 26; Gaps 6;

Qy 9 MRAAALAAVAVLVSSCGKGDAAOGOPAGREAPAPVGVVTVHQTVALTVTELGR 67
 Db 5 LRTFGLCAITVAALAAACK-----PDQQAAPPEPEVSVLEMKPTTLEEDLVGR 55
 Qy 68 ESLRTADVRAOYGGIIQKRLFOEGSYVRAQPLYOIDSSTYEANLESARALATAQAT 127
 Db 56 SAYASADVRAVDEVLKRLYTEGANTYEGQPLFOIDPSQKATLLQAQGLAAEAATYT 115
 Qy 128 KADADLARYKPLVAEAVSRQEDYDAVYAKSAAGVAAQAIAKAGININRSRTAP 187
 Db 116 NAKIAAARARSIALPQOYVSRAIDITAEANERSGANYQAARGAVEAARIQLGFATV 175
 Qy 188 SGFTGSKVSEGLTLNAGDTTVLATIRONTPMYVNVVTVOSASEVMKLRQIAEGKL 246
 Db 176 TGRAGIQRVTEGALVAGEATLLTTVDQIDPLVNPMSSELAALQAQSSGNVQLSGD 235
 Qy 247 GVIAVGIKFDGTVYPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYVRV 306
 Db 236 GKSSINVELGNGTQYOHPTGLDVSAYVYDPSTGAVSLRATLPNPGSLIPRAFTFR 295
 Qy 307 VAVDNAFVVPQAAVTRGAKDT-VMI VNAOGMEPREVTVAOOGGTNMI VTSGLK 365
 Db 296 GORNNAYLLPQAAVQORATGAYALVLGKDGKVRKNLITVDQQKQGWIVTGKMT 355

Qy 366 VEGISIAITGAKKVTPEK-----ASSENQAAAPOSGVOTASEAKTASEA 411
 Db 356 VDGVRKAKKEGQPAKGVF--WDPNKPAQGGQGAAGAAPAAAGDGAQAPADRADAA 411

RESULT 40

Q93PUS PRELIMINARY; PRT; 391 AA.
 ID Q93PUS
 AC Q93PUS;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Periplasmic efflux pump component Ttgg.
 GN TTGG.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 CX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DOT-TIE;
 RX MEDLINE=21289082; PubMed=11395460;
 RA Rojas A., Duque E., Mosqueda G., Golden G., Hurtado A., Ramos J.L.,
 RA Segura A.;
 RT "Three Efflux Pumps Are Required To Provide Efficient Tolerance to
 RT Toluene in Pseudomonas putida DOT-TIE";
 RL J. Bacteriol. 183:3967-3973(2001).
 DR EMBL; AF299253; AA069563.1; -;
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 SQ SEQUENCE 391 AA; 4258 MW; E2362606D71E52A1 CRC64;

Query Match 31.3%; Score 631; DB 2; Length 391;
 Best Local Similarity 35.5%; Pred. No. 3.9e-27;
 Matches 141; Conservative 84; Mismatches 140; Indels 32; Gaps 8;

Qy 7 KAMRAALAAVAVLVSSCGKGDAAOGOPAGREAPAPVGVVTVHQTVALTVTELGR 66
 Db 17 RALRVPLTA--LMLISGCE-----KEQVSATPPDPVGVTVYAAQALTTTDLRGR 67
 Qy 67 LESLRTADVRAOYGGIIQKRLFOEGSYVRAQPLYOIDSSTYEANLESARALATAQAT 126
 Db 68 TSAFRAVEVAPQVSGILQKRSFVEGAHVKLQGLQYQIDPRTYEQLRAEANRRTSAQ-NL 126
 Qy 127 AKADADLARYKPLVAEAVSRQEDYDAVYAKSAAGVAAQAIAKAGININRSRTAP 186
 Db 127 AR-----RYETLLKTKAVSKQOYDDALAAMKQAEADYQVAR-----IDVYTRV 173
 Qy 187 ISGFTGSKVSEGLTLNAGDTTVLATIRONTPMYVNVVTVOSASEVMKLRQIAEGKL 245
 Db 174 ISGRIGSTVTEGALVNGAQSIAITVTDPIYVDTOPIYTKLGLHKLBSGRLOK 233
 Qy 246 DGVIANGIKFDGTVYPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYVRV 305
 Db 234 ENQAEVSLTIDDSAVPLPGLTKFSEVSCDPTGVSYTLRAEPNPNKLLPGMVFHLLK 293
 Qy 306 QVAVDNAFVVPQAAVTRGAK--DTVMI VNAOGMEPREVTVAOOGGTNMI VTSGLK 363
 Db 294 EGVQNAAILVPOQAISHDRTGVPVSWVMKADNVESEIDTTLRTVGNAMLSNGVTGER 353
 Qy 364 VVEGIS-----IAGITGAKKVTPEKMASSSENQAAA 394
 Db 354 IITEGVORVRSIGIAVNAVEAKVNLVDPGFAATTEASA 390

Search completed: September 8, 2003, 14:02:11
 Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:59:31 / Search time 21 Seconds
(without alignments)
1886.738 Million cell updates/sec

Title: US-09-889-756A-2
Perfect score: 2019
Sequence: 1 MAFYAFKRAMRAALAAVAL.....AAPQSGVQTASEAKTASAE 412

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1993	98.7	412	2	H81825
2	1989	98.5	412	2	H81051
3	1278	63.3	271	2	S42418
4	770	38.1	388	2	H87349
5	762.5	37.8	397	2	D90693
6	762.5	37.8	397	2	H85543
7	762.5	37.8	397	2	A36938
8	756	37.4	397	2	AG0561
9	753	37.3	383	2	G39629
10	750.5	37.2	385	2	G65119
11	749.5	37.1	385	2	A91146
12	749.5	37.1	385	2	H85991
13	726.5	36.0	395	2	AD0380
14	723	35.8	404	2	AE2950
15	723	35.8	427	2	F98332
16	699.5	34.6	387	2	A83072
17	697.5	34.5	373	2	H85757
18	697.5	34.5	373	2	G90861
19	670	33.2	408	2	H82600
20	656	32.5	385	2	S47723
21	655	32.4	385	2	A91178
22	655	32.4	385	2	B86024
23	628.5	31.1	386	2	AD0423
24	628	31.1	396	2	H83393
25	595	29.5	400	2	AE2889
26	595	29.5	410	2	B97665
27	537	26.6	395	2	AH3455
28	482	23.9	388	2	G95375
29	460	22.8	414	2	H83335

30	459.5	22.8	373	2	G96007	probable acriflavi
31	446	22.1	414	2	T30829	hypothetical prote
32	442.5	21.9	396	2	AG2881	HlyD family secret
33	442.5	21.9	396	2	G97657	hypothetical prote
34	432	21.4	464	2	E85834	probable membrane
35	432	21.4	464	2	A64974	hypothetical prote
36	431	21.3	413	2	AC0771	probable membrane
37	431	21.3	464	2	B90989	probable membrane
38	430	21.3	426	2	A83330	probable RND efflu
39	426.5	21.1	451	2	A13623	acriflavin resista
40	425.5	21.1	444	2	AH0346	putative HlyD fami
41	424.5	21.0	367	2	E81379	probable membrane
42	403.5	20.0	425	2	AH0122	probable drug effl
43	394	19.5	385	2	A83466	probable RND efflu
44	387	19.2	411	2	A82566	membrane fusion pr
45	356	17.6	425	2	AC2959	HlyD family secret

ALIGNMENTS

RESULT 1									
H81825									
membrane fusion protein NMA1970 [imported] - Neisseria meningitidis (strain Z2491 serogr									
C/Species: Neisseria meningitidis									
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001									
C/Accession: H81825									
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell									
; Holtz, S.; Jørgensen, K.; Leather, S.; Mouton, S.; Mungall, K.; Quail, M.A.; Rajandream,									
Nature 404, 502-506, 2000									
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.									
A/Reference number: AB1775; MUID:20222556; PMID:10761919									
A/Accession: H81825									
A>Status: Preliminary									
A/Molecule type: DNA									
A/Residues: 1-412 <PAR>									
A/Cross-references: GB:AL162757; GB:AL157959; NID:97380371; PIDN:CAB85190.1; PID:9738060.									
A/Experimental source: serogroup A, strain Z2491									
C/Genetics:									
A/Gene: mtrC; NMA1970									
Query Match									
Best Local Similarity 98.7%; Score 1993; DB 2; Length 412;									
Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;									
QY	1	MAFYAFKRAMRAALAAVALIVSSCGKGDPAAGGQPAAGREAPAVGVVTHPQVALT	60						
DB	1	MAFYAFKRAMRAALAAVALIVSSCGKGDPAAGGQPAAGREAPAVGVVTHPQVALT	60						
QY	61	VEI.PGRLESIRTDVRAOVGGIIQKRLFORGSVVRAGQPIYOTIDSSTYEANTLSARQOLA	120						
DB	61	VEI.PGRLESIRTDVRAOVGGIIQKRLFORGSVVRAGQPIYOTIDSSTYEANTLSARQOLA	120						
QY	121	TAQATLAKADADARVYPLVAEAVEROEYDAVATKASAEVKAQAQAIKSAGINLR	180						
DB	121	TAQATLAKADADARVYPLVAEAVEROEYDAVATKASAEVKAQAQAIKSAGINLR	180						
QY	181	SRTTAPISGFTGSKVSEGLTLNAGDTTVIATIRQNPVNVNVTQSASEVMKLRQIAEG	240						
DB	181	SRTTAPISGFTGSKVSEGLTLNAGDTTVIATIRQNPVNVNVTQSASEVMKLRQIAEG	240						
QY	241	KLLAADGVAVGKPDGTVYPERGRLFPADPVNNSGQITLRAVPNDQNLMLGLVY	300						
DB	241	KLLAADGVAVGKPDGTVYPERGRLFPADPVNNSGQITLRAVPNDQNLMLGLVY	300						
QY	301	RVLMDQVAVDNAFVVPQAVTRGAKDTVMIVNAQGMREPREVVAQOQGNMTVTSGLKD	360						
DB	301	RVLMDQVAVDNAFVVPQAVTRGAKDTVMIVNAQGMREPREVVAQOQGNMTVTSGLKD	360						
QY	361	GDKRVVEGISIAGITGAKVTPKEMASSENQAAPQSGVQTASEAKTASAE	412						
DB	361	GDKRVVEGISIAGITGAKVTPKEMASSENQAAPQSGVQTASEAKTASAE	412						

RESULT 2
F81051
membrane fusion protein NMB1716 [imported] - Neisseria meningitidis (strain MC58 serogroup C)
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jun-2001
C:Accession: F81051
C:Title: H. J. Saunders, N. J. Heidelberg, J. J. Jeffries, A. C. Nelson, K. E. Eisen, J. A. Hickey, E. K. Haft, D. H. Salzberg, S. L. White, O. J. Fleischmann, R. D. Dougherty, B. A. H. H. Qiu, H. Vamathevan, J. J. Gill, J. Scarlato, V. Masignani, V. Pizze, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H. O.; Fraser, C. M.; Moxon, E. R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: F81051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <TEXT>
A:Cross-references: GB:A8002521; GB:A8002098; NID:g7226962; PIND:AAK2063.1; PID:g722697
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1716

Query Match 98.5%; Score 1989; DB 2; Length 412;
Best Local Similarity 98.5%; Pred. No. 6,4e-114;
Matches 406; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAFYAFKAMRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60
DB 1 MAFYAFKAMRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60

QY 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120
DB 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120

QY 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLR 180
DB 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLR 180

QY 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLR 180
DB 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLR 180

QY 181 SRITAPISGFIGSGKVSSEGTILNAGDTVTATIRQTNPMYVNTQASAEVWKLRLRQIAEG 240
DB 181 SRITAPISGFIGSGKVSSEGTILNAGDTVTATIRQTNPMYVNTQASAEVWKLRLRQIAEG 240

QY 241 KLLAADGVIAVGIKFDDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDQNIIMPGLYV 300
DB 241 KLLAADGVIAVGIKFDDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDQNIIMPGLYV 300

QY 301 RVLMDQVAVDNAFVVPQOAVTRGAKDTWIVNACGMEPREVTVAQOOGTWTWISGLKD 360
DB 301 RVLMDQVAVDNAFVVPQOAVTRGAKDTWIVNACGMEPREVTVAQOOGTWTWISGLKD 360

QY 361 GDKVNVGISTAGTIGAKKVTPEKMASSENOAAPOSGVOTRASEAKTASEAK 412
DB 361 GDKVNVGISTAGTIGAKKVTPEKMASSENOAAPOSGVOTRASEAKTASEAK 412

RESULT 3
S42418
probable antibiotic resistance protein mtrC - Neisseria gonorrhoeae (strain CH95) (fragm
C:Species: Neisseria gonorrhoeae
C>Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 08-Oct-1999
C:Accession: S42418; S40252
C:Title: R. Pan, W. Spratt, B. G. Mol. Microbiol. 11, 769-775, 1994
A:Title: Regulation of the permeability of the gonococcal cell envelope by the mtr syste
A:Reference number: S42417; MUID:94254732; PMID:8196548
A:Accession: S42418
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-271 <PAN>
A:Cross-references: EMBL:Z25796; NID:g438190; PIND:CAA81046.1; PID:g438192
A:Experimental source: CH95
C:Genetics:

A:Gene: mtrC
C:Superfamily: 1:ipoy1/biotin-binding homology
F:68-111,183-211/Domain: 1:ipoy1/biotin-binding homology #status atypical <LPB>

Query Match 63.3%; Score 1278; DB 2; Length 271;
Best Local Similarity 96.7%; Pred. No. 7.1e-71;
Matches 262; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAFYAFKAMRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60
DB 1 MAFYAFKAMRAAALAAVALVSSCGKGDAAGGAPAGREAPAPVGVVTHPQTVALT 60

QY 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120
DB 61 VELPGRLESRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 120

QY 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLR 180
DB 121 TAOATLAKADADLARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLR 180

QY 181 SRITAPISGFIGSGKVSSEGTILNAGDTVTATIRQTNPMYVNTQASAEVWKLRLRQIAEG 240
DB 181 SRITAPISGFIGSGKVSSEGTILNAGDTVTATIRQTNPMYVNTQASAEVWKLRLRQIAEG 240

QY 241 KLLAADGVIAVGIKFDDGTVPYPEKGRLLFAD 271
DB 241 KLLAADGVIAVGIKFDDGTVPYPEKGRLLFAD 271

RESULT 4
E87349
HlyD family secretion protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87349
C:Title: R. Nierman, W. C. J. Feldblyum, T. V. Paulsen, I. T. Nelson, K. E. Eisen, J. J. Heidelberg, J. J. B. J. Laub, M. T. DeBoy, R. T. Dodson, R. J. Durkin, A. S. Gwinn, M. L. Haft, D. H. Kolon
n. J. Ermolaeva, M. J. White, O. J. Salzberg, S. L. White, L. J. Shapiro, J. C. J. Fraser, C. M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <STO>
A:Cross-references: GB:A8005673; NID:g13422055; PIND:AAK22793.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0808

Query Match 38.1%; Score 770; DB 2; Length 388;
Best Local Similarity 46.2%; Pred. No. 9.5e-40;
Matches 180; Conservative 57; Mismatches 11; Indels 22; Gaps 9;

QY 16 AAVALVSSCG-KGGDAAGGAPAGREAPAPVGVVTHPQTVALTYELPGRLESRTAD 74
DB 15 AAVALVSSCG-KGGDAAGGAPAGREAPAPVGVVTHPQTVALTYELPGRLESRTAD 74

QY 75 VRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 132
DB 75 VRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 132

QY 133 LARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLRSRITAPISGFIG 192
DB 133 LARYPLVAEAVSRQBYDAATVAKSAEAGVAAQAQAIKSAGINLRSRITAPISGFIG 192

QY 193 QSKVSEGLTLAGDTVTATIRQTNPMYVNTQASAEVWKLRLRQIAEGKLLAADGVIAVG 252
DB 193 QSKVSEGLTLAGDTVTATIRQTNPMYVNTQASAEVWKLRLRQIAEGKLLAADGVIAVG 252

QY 253 IKPDDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDQNIIMPGLYVRLMDQVAVDNA 312
DB 253 IKPDDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDQNIIMPGLYVRLMDQVAVDNA 312

QY 312 LKLEDSGYPIPGRLSRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 360
DB 312 LKLEDSGYPIPGRLSRTADVRAOVGGIIQKRLFOEGSYVRACOPLYQIDSTYEANLESARAQLA 360

Qy 313 FVVPQOAVTRGAKD--TWIVVAAGGMEPREVTVAOQOGTMMVTSGLKDDGVVVEGIS 370
 Db 307 ILIPQAVNRDPKPGKATVWLVAOK-GPEPRPVTLGQTVGDKMLVTSGLNAGDKVIVGEL- 364
 Qy 371 IAGITGAKKVTPEKVASSENQAAAPQSGVQ 400
 Db 365 -----MKVPRGAPITKAVPRAGAAPAAQAQ 387

RESULT 5

acridine efflux pump [imported] - Escherichia coli (strain O157:H7, substrain RMD 05095
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C/Accession: D90693
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: D90693
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-397 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA833939.1; PID:G13359973; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RMD 050952
 C/Genetics:
 A/Gene: EC60516

Query Match 37.8%; Score 762.5; DB 2; Length 397;
 Best Local Similarity 42.2%; Pred. No. 2.8e-39;
 Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

Qy 11 AALAAVAALVSSCGKGGDAAGGAPAGREAPVVGVTVAHQVATVTELPGRLESL 70
 Db 11 AVVLMSSGLATGTC-DDKQAQGGQ-----QMPAVGVVTKTEPLQITTELPGRISAY 63
 Qy 71 RTADVRAQVGGIIOKRLFOESVYRAGQPLYOIDSSTEANLESARQALATQATLAKAD 130
 Db 64 RIAVRRPQVSGIILKRNFKESGDIKAGVSLVQIDPATYQATYDAKGLAKAQAANITQ 123
 Qy 131 ADLARYPELVAAEAVSRQEDYDAVTAKRSAAGVKAQAATKAGININRSRTAPISGF 190
 Db 124 LTVNRVQQLGTQYTSKQEDQALADAAQANAATAAQAETARINLAATKVTSPISGR 183
 Qy 191 IGGKSVSGTLLNAGDTTVAITRTQNMVYNTQSASEWKLKROIAEGTLAADGVIA 250
 Db 184 IGSNVTGALVQNGQATATATVQQLDPIYDVTOSSNDPLRLKQELANGTLKQENGKAK 243
 Qy 251 VGIKFDGTVVPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYRVLMDOYAVD 310
 Db 244 VSLTSDGIRKFPQDGTLEFSDVTVDTGTSITLRAIFNPHTLLPGMFVAREBEGINP 303
 Qy 311 NAFVPPQOAVTRGAKD--TWIVVAAGGMEPREVTVAOQOGTMMVTSGLKDDGVVVEG 368
 Db 304 NALIVPQOGVTRPRGATVAVGADDKVETRPVISAQIDKMLVTEGLAAGRVVVISG 363
 Qy 369 ISLIGITGAKKVT-----KEMASSENQAAAPQSGVQ 400
 Db 364 L-----QKVRGQVQKAEVTAADNNQAA--SGAQ 391

RESULT 6

H85543
 acridine efflux pump [imported] - Escherichia coli (strain O157:H7, substrain EDJ933)
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: H85543
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: H85543
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-397 <STD>
 A/Cross-references: GB:AE005174; NID:G12513330; PIDN:AA654812.1; GSPDB:GN00145; UMGCP:205
 A/Experimental source: strain O157:H7, substrain EDJ933
 C/Genetics:
 A/Gene: acrA

Query Match 37.8%; Score 762.5; DB 2; Length 397;
 Best Local Similarity 42.2%; Pred. No. 2.8e-39;
 Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

Qy 11 AALAAVAALVSSCGKGGDAAGGAPAGREAPVVGVTVAHQVATVTELPGRLESL 70
 Db 11 AVVLMSSGLATGTC-DDKQAQGGQ-----QMPAVGVVTKTEPLQITTELPGRISAY 63
 Qy 71 RTADVRAQVGGIIOKRLFOESVYRAGQPLYOIDSSTEANLESARQALATQATLAKAD 130
 Db 64 RIAVRRPQVSGIILKRNFKESGDIKAGVSLVQIDPATYQATYDAKGLAKAQAANITQ 123
 Qy 131 ADLARYPELVAAEAVSRQEDYDAVTAKRSAAGVKAQAATKAGININRSRTAPISGF 190
 Db 124 LTVNRVQQLGTQYTSKQEDQALADAAQANAATAAQAETARINLAATKVTSPISGR 183
 Qy 191 IGGKSVSGTLLNAGDTTVAITRTQNMVYNTQSASEWKLKROIAEGTLAADGVIA 250
 Db 184 IGSNVTGALVQNGQATATATVQQLDPIYDVTOSSNDPLRLKQELANGTLKQENGKAK 243
 Qy 251 VGIKFDGTVVPEKGRLLFADPVVNESTGQITLRAAVNDONILMPGLYRVLMDOYAVD 310
 Db 244 VSLTSDGIRKFPQDGTLEFSDVTVDTGTSITLRAIFNPHTLLPGMFVAREBEGINP 303
 Qy 311 NAFVPPQOAVTRGAKD--TWIVVAAGGMEPREVTVAOQOGTMMVTSGLKDDGVVVEG 368
 Db 304 NALIVPQOGVTRPRGATVAVGADDKVETRPVISAQIDKMLVTEGLAAGRVVVISG 363
 Qy 369 ISLIGITGAKKVT-----KEMASSENQAAAPQSGVQ 400
 Db 364 L-----QKVRGQVQKAEVTAADNNQAA--SGAQ 391

RESULT 7

A36938
 acriflavin resistance protein acrA precursor - Escherichia coli (strain K-12)
 C/Species: Escherichia coli
 C/Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 01-Mar-2002
 C/Accession: A36938; F64776
 R/Ma, D.; Cook, D.N.; Alberti, M.; Pon, N.G.; Nikaido, H.; Hearst, J.E.
 J. Bacteriol. 175, 6299-6313, 1993
 A/Title: Molecular cloning and characterization of acrA and acrE genes of Escherichia co
 A/Reference number: A36938; MUID:94012493; PMID:8407802
 A/Accession: A36938
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-397 <MAA>
 A/Cross-references: GB:U00734; NID:G392830; PIDN:AA67134.1; PID:G5523310
 R/Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: F64776
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-397 <BAT>
 A/Cross-references: GB:AE000152; GB:U00096; NID:G1786660; PIDN:AACT3565.1; PID:G1786668;
 A/Experimental source: strain K-12, substrain MG1655
 C/Genetics:
 A/Gene: acrA
 A/Function:
 A/Description: multidrug efflux pump
 A/Note: active form contains proteins acrA and acrB

C:Superfamily: lipoyl/biotin-binding homology

C:Keywords: lipoprotein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-397/Product: acriflavin resistance protein acra #status predicted <MAT>

F:61-104,176-205/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 37.8%; Score 762.5; DB 2; Length 397;

Best Local Similarity 42.2%; Pred. No. 2,8e-39;

Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

```

Oy 11 AALAAALVALVSSCGKGDAAGGQAPAGREAPAVGVVTHPQTVALTVLPEGRLESL 70
Db 11 AVLMISGLSLATGCGD-DDKQAQGGQ-----QMAVGVVTVKTEPLDITTELPRTSAY 63
Oy 71 RTADVRAVGGIIGRLFOEGSVYRAGPLVQIDISTEAMNESARAQATATLAKAD 130
Db 64 RIAEVRPVSGLIKRNFKEGSDIFAGVSLVQIDPATYQATYDSAKGLARQAAMANIQ 123
Oy 131 ADLARYPELVAAEAVSRQEDAAVTAKRSAGGVAKQAQAIKSAGININRSRTAPISGF 190
Db 124 LTVNRYSKGLTGYISKQEDYDALADQAQANAAVTAAKAAVETARINLATYKVTSPISGR 183
Oy 191 IGKSVSSEGLTNAGDTTVLATIRQTNPMYVNVTSASEWKLRRQIAEGKILAADGVIA 250
Db 184 IGKSVTEGALVQNGAALATVQQLDPIYDVTSNDPRLKQELANGTLKQENGKAK 243
Oy 251 VGIFDDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDONILMPGLYVRVLMDOVAVD 310
Db 244 VSLTSDGIKFPQDGTLEFSDVTYVQTGISTTLRAIFPNPDHTLLPGFVPAARLEGINP 303
Oy 311 NAFVVPQQAATVRGAKD--TWMIVNAQGMPEREVTVAAQOQNTMIVTSGLKDGKRVVVG 368
Db 304 NALIVPQGGVTRPRGDAFTLVVGADKVFETRPVIAVSAQIGDKMLVTGLKAGDREVVISG 363
Oy 369 ISIGITGAKKVT-----KEMASSENQAAPQSGVQ 400
Db 364 L-----QKRPAGVQVKAQEVITADNNQQA--SGAQ 391

```

RESULT 8

AG0561 acriflavin resistance protein A precursor [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AG0561

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, C.; Mouton, S.; O'Garra, P.

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AG0561

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-397 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD04961.1; PID:q16501748; GSPDB:GN00176

C:Genetics:

A:Gene: STY0520

Query Match 37.4%; Score 756; DB 2; Length 397;

Best Local Similarity 41.3%; Pred. No. 7e-39;

Matches 166; Conservative 78; Mismatches 129; Indels 30; Gaps 5;

```

Oy 11 AALAAALVALVSSCGKGDAAGGQAPAGREAPAVGVVTHPQTVALTVLPEGRLESL 70
Db 11 AVLMISGLSLATGCGD-DDKQAQGGQ-----QMEVGVVTVKTEPLDITTELPRTVAY 63
Oy 71 RTADVRAVGGIIGRLFOEGSVYRAGPLVQIDISTEAMNESARAQATATLAKAD 130
Db 64 RIAEVRPVSGLIKRNFKEGSDIFAGVSLVQIDPATYQATYDSAKGLARQAAMANI 123

```

```

Oy 131 ADLARYPELVAAEAVSRQEDAAVTAKRSAGGVAKQAQAIKSAGININRSRTAPISGF 190
Db 124 LTVNRYSKGLTGYISKQEDYDALADQAQANAAVTAAKAAVETARINLATYKVTSPISGR 183
Oy 191 IGKSVSSEGLTNAGDTTVLATIRQTNPMYVNVTSASEWKLRRQIAEGKILAADGVIA 250
Db 184 IGKSVTEGALVQNGAALATVQQLDPIYDVTSNDPRLKQELANGTLKQENGKAK 243
Oy 251 VGIFDDGTVPYPEKGRLLFADPVNVESTGOITLRAAVNDONILMPGLYVRVLMDOVAVD 310
Db 244 VSLTSDGIKFPQDGTLEFSDVTYVQTGISTTLRAIFPNPDHTLLPGFVPAARLEGINP 303
Oy 311 NAFVVPQQAATVRGAKD--TWMIVNAQGMPEREVTVAAQOQNTMIVTSGLKDGKRVVVG 368
Db 304 NALIVPQGGVTRPRGDAFTLVVGADKVFETRPVIAVSAQIGDKMLVTGLKAGDREVVISG 363
Oy 369 ISIGITGAKKVT-----KEMASSENQAAPQSGVQ 397
Db 364 L-----QKRPAGVQVKAQEVITADNNQQA--SGAQ 391

```

RESULT 9

multidrug-efflux transport protein A precursor - Pseudomonas aeruginosa

N:Alternate names: multidrug resistance protein A

C:Species: Pseudomonas aeruginosa

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Dec-2000

C:Accession: S39629; D83593

R:Pooler, K.; Heinrichs, D.E.; Neshat, S.

Mol. Microbiol. 10, 529-544, 1993

A:Title: Cloning and sequence analysis of an EnvC homologue in Pseudomonas aeruginosa

A:Reference number: S39629; MUID:95058196; PMID:7968531

A:Accession: S39629

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-383 <POO>

A:Cross-references: GB:L11616; NID:9438852; PIDN:AAA74436.1; PID:9438853

A:Experimental source: strain CD10

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: D83593

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-383 <STO>

A:Cross-references: GB:AE004479; GB:AE004091; NID:9946272; PIDN:AA03814.1; GSPDB:GN00176

A:Experimental source: strain PA01

C:Genetics:

A:Gene: mexA; PA0425

C:Function:

A:Description: probably involved in secretion of the siderophore pyoverdine

A:Note: inducible under conditions of iron limitation

C:Superfamily: lipoyl/biotin-binding homology

C:Keywords: lipoprotein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-383/Product: multidrug-efflux transport protein A #status predicted <MAT>

F:59-102,160-188/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 37.3%; Score 753; DB 2; Length 383;

Best Local Similarity 42.1%; Pred. No. 1e-38;

Matches 174; Conservative 66; Mismatches 125; Indels 46; Gaps 10;

```

Oy 8 AMRAALAAVAL-VLSSCGKGDAAGGQAPAGREAPV-----GVVTVHPQTVALTVE 62
Db 6 AMRLVLPALVAISLGGCKS-----EAPPPAQTEPVGIVTLEAQTIVLNTTE 53
Oy 63 LPRLESLRTADVAVQVGGIIGRLFOEGSVYRAGPLVQIDISTEAMNESARAQATATA 122
Db 54 LPRITAFRIAEVAPVQVGGIIGRLFKESGDVAKAQQLVQIDPATYEAQVQSAQANLAST 113

```


QY 123 QATLAKADADLARKYPLVAEAVSRQEDAAVTAKRSAGVKAQAQAIKASGINLNRSR 182
 Db 114 Q-----EQARQKYLIVADQAVSKQY-----ADANAAVILQSKAVEQARINLERYT 159
 QY 183 ITAPISGPIGOSKVSSEGLTLNAGDPTVATIROTNPMVYVNTOSASEVMKLRQIAGSKL 242
 Db 160 VLSISGRIGSVAETEGALVTNNGQANNAVIVQQLDPIYDVTQSPSTALLRRRLASGQL 219
 QY 243 -LAADGVLAVGIRKEDDGTVPYPEKGRLLFADPVNVESTGQITLRAAVPNDQVILMPGLYVR 301
 Db 220 ERADBNNAKYSKLKEDSGQYPLBGRLESEVSDEGTSVITIRAFVFNPNELLPGMFVH 279
 QY 302 VLMDQVAVDNAFVVPQQAATRGAKD--TVMIVNAQGMEPREVTVAQOQGTNMTVITSGLK 359
 Db 280 AOLQEGVAKIILAPPOGQVTRDLKGQATALVYVNAQNKVELKVIKADRVIGDKMLVTGSLN 339
 QY 360 DGDVYVVEGTSIA--GITGAKKVTPEKMASSENQAAAPQSGVGTASEAKTASE 410
 Db 340 AGDKITIEGLQFVQPGVE-VKTVPAKNVSAQAKADAP-----AKTDSK 382

RESULT 10

C65119

acriflavin resistance protein acrf precursor - Escherichia coli (strain K-12)
 N/Alternate names: envelope protein
 C/Species: Escherichia coli
 C/Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 01-Mar-2002
 C/Accession: C65119; S18536; S18665
 R/Biatcher, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
 A.; Rose, D.; Mau, B.; Siao, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: C65119
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-385 <BIAT>
 A/Cross-references: GB:AE000405; GB:U00096; NID:G1789659; PIDN:AACT6297.1; PID:G1789665;
 A/Experimental source: strain K-12, substrain MG1655
 R/Klein, J.R.; Henrich, B.; Plapp, R.
 Mol. Gen. Genet. 230, 230-240, 1991
 A/Title: Molecular analysis and nucleotide sequence of the envC operon of Escherichia C
 A/Reference number: S18536; MUID:92079901; PMID:1720861
 A/Accession: S18536
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-190, 'AL', 193, 'SLW', 197, 'KRLNWRILSSSILST', 213-312, 'T', 315-385 <KLE>
 A/Cross-references: EMBL:X57948
 R/Klein, J.R.; Henrich, B.; Plapp, R.
 Curr. Microbiol. 21, 341-347, 1990
 A/Title: Molecular cloning of the EnvC gene of Escherichia coli.
 A/Reference number: S18665
 A/Accession: S18665
 A/Status: preliminary; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-96 <KLE>
 A/Cross-references: EMBL:X57948; NID:9510827; PIDN:CAA41016.1; PID:9510829
 A/Experimental source: strain K-12
 C/Genetics:
 A/Gene: acrf, envC
 C/Superfamily: lipoyl/biotin-binding homology
 C/Keywords: cell division; lipoprotein
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-385/Product: acriflavin resistance protein acrf #status predicted <MAT>
 F:60-103, 175-204/Domain: lipoyl/biotin-binding homology #status atypical <LBP>

Query Match 37.1%; Score 750.5; DB 2; Length 385;
 Best Local Similarity 42.1%; Pred. No. 1.4e-38;
 Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

QY 17 AVALVSSCGKGDAAGQGPAGREAPVVGVTVA--PQTVALTV--ELPGRLESIRTA 73
 Db 17 SAALINGCNDKGEKAVHGEPO-----VTVAIVTAPLEVTLPGRTNAYRIA 65

QY 74 DVAQVGGIIQKRLFOEGSVYRAGQPLVYQIDSSSTEANLESARAQLATAQATLAKADAL 133
 Db 66 EVRPQVSGIYLVNNTFEGSDVQAGQSLYQIDPATYQANVYSAKELAKSEAAAIHLTV 125
 QY 134 ARYKPLVAEAVSRQEDAAVTAKRSAGVKAQAQAIKASGINLNRSRTTAPISGTIGQ 193
 Db 126 KRYVPLVGTXYISQOEYDAIADARQADAIVIAKATVESARINLAATYKTAPISGRIGK 185
 QY 194 SKVSEGLTLNAGDPTVATIROTNPMVYVNTOSASEVMKLRQIAGSKLAAQGVAVGI 253
 Db 186 STYEGALVTNNGQTTTELATVQQLDPIYDVTQSSNDPRLKQSVQGNLHKENTSNVEL 245
 QY 254 KPDDGTVPYPEKGRLLFADPVNVESTGQITLRAAVPNDQVILMPGLYVRVLMDQVAVDNAF 313
 Db 246 VMENGTPTPLKGLTQFSVDVTVDESTGSLIRAFVFNPNQHTLLPGMFARIDEGVQDAI 305
 QY 314 VVPQQAATRGAKD--TVMIVNAQGMEPREVTVAQOQGTNMTVITSGLKDGKRVVVEGISI 371
 Db 306 LIPOGVSRTPRGDATVLIVNDKSQVEARPVVASQAIGDKMLISEGKSGDQVIVSGL-- 363
 QY 372 AGITGAKKVTPEKMASSENQAAAPQSGVGTASEAKTASE 410
 Db 364 -----QKARPG-----QVKATTDTPTADTASK 385

RESULT 11

A91146

hypothetical protein Ecs4137 [imported] - Escherichia coli (strain O157:H7, substrain R1M
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C/Accession: A91146
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: A91146
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-385 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA837560.1; PID:G13363610; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:
 A/Gene: Ecs4137

Query Match 37.1%; Score 749.5; DB 2; Length 385;
 Best Local Similarity 42.1%; Pred. No. 1.7e-38;
 Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;

QY 17 AVALVSSCGKGDAAGQGPAGREAPVVGVTVA--PQTVALTV--ELPGRLESIRTA 73
 Db 17 SAALINGCNDKGEKAVHGEPO-----VTVAIVTAPLEVTLPGRTNAYRIA 65
 QY 74 DVAQVGGIIQKRLFOEGSVYRAGQPLVYQIDSSSTEANLESARAQLATAQATLAKADAL 133
 Db 66 EVRPQVSGIYLVNNTFEGSDVQAGQSLYQIDPATYQANVYSAKELAKSEAAAIHLTV 125
 QY 134 ARYKPLVAEAVSRQEDAAVTAKRSAGVKAQAQAIKASGINLNRSRTTAPISGTIGQ 193
 Db 126 KRYVPLVGTXYISQOEYDAIADARQADAIVIAKATVESARINLAATYKTAPISGRIGK 185
 QY 194 SKVSEGLTLNAGDPTVATIROTNPMVYVNTOSASEVMKLRQIAGSKLAAQGVAVGI 253
 Db 186 STYEGALVTNNGQTTTELATVQQLDPIYDVTQSSNDPRLKQSVQGNLHKENTSNVEL 245
 QY 254 KPDDGTVPYPEKGRLLFADPVNVESTGQITLRAAVPNDQVILMPGLYVRVLMDQVAVDNAF 313
 Db 246 VMENGTPTPLKGLTQFSVDVTVDESTGSLIRAFVFNPNQHTLLPGMFARIDEGVQDAI 305
 QY 314 VVPQQAATRGAKD--TVMIVNAQGMEPREVTVAQOQGTNMTVITSGLKDGKRVVVEGISI 371
 Db 306 LIPOGVSRTPRGDATVLIVNDKSQVEARPVVASQAIGDKMLISEGKSGDQVIVSGL-- 363

Db 60 TRLAEVRPRVSGIIIVRVEGSGLVKEGDLVLRIDRAPFOVRVDSAGETLLRRAQAOLQA 119

QY 130 DADLARYKPLVAEAASVROEYDAAVTAKRSAAAGVKAQAQAIAKSAGINLNRSRTAPISG 189

Db 120 RQTADROOQLRARSVNGSQGFEDNAILLAQADAEVAVAEAGVABERLNLYADVKAPISG 179

QY 190 FIGOSKVSSEGLTLNAGDTTLATIRQTNPMYVNTQSAEVMKLRRQIAESKLLAADGVI 249

Db 180 VIGRARITEGALVATSGSENLATIQQLDPIYADFQPADILRLRKALQDQGLMTGQNEA 239

QY 250 AVGKPEDDGTVPYPERKGLFPADPVNVESTGOITLRAVPNONTIMLGLYRVLMDQYAV 309

Db 240 EVNLFPDGSRYRPSVGRLLPSEAAVDETTGQVTLRGEPFNPNGDLLPQMYRVQIQQGIQ 299

QY 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPREVTVAQOQGTWMTVTSGLKDGDKYVE 367

Db 300 KAAFAVVPQQAQVORAGGASVLYVNAEDTVEQRVSVRSIGDKWVISEGLDDGDRYVAE 359

QY 368 GISIAGITGAKKVTPEKEMASSENQAAPQSGVQTASEAKTASE 410

Db 360 GFOKTA-PGA-KVKPEPMSQEPDVAAA--AGSEGAAPSETSSSE 398

RESULT 15

F98332
hypochemical protein AGR_L_3215 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: F98332
R/Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurrello, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: F98332
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-427 <KTR>
A/Cross-references: GB:AE007870; PIDN:AAK90184.1; PID:g15160189; GSPDB:GN00170
C/Genetics:
A/Gene: AGR_L_3215
A/Map position: linear chromosome

Query Match 35.8%; Score 723; DB 2; Length 427;
Best Local Similarity 40.0%; Pred. No. 7.8e-37;
Matches 161; Conservative 83; Mismatches 143; Indels 16; Gaps 7;

QY 11 AALAAALVALVSSCGKGDAAQGGQAGREAPAP-VGVVTVHPQTVALTELPGRLS 69

Db 32 AAVLLTGLIV-----GCSDEQASADA--APPGAIVKVAVKPEELPTINELPGRIAP 82

QY 70 LRTADVPAQVGGITQKRLFOEGSVYRAGQPLXOIDSSTYEANLESARQAOLATQAATLAKA 129

Db 83 TRLAEVRPRVSGIIIVRVEGSGLVKEGDLVLRIDRAPFOVRVDSAGETLLRRAQAOLQA 142

QY 130 DADLARYKPLVAEAASVROEYDAAVTAKRSAAAGVKAQAQAIAKSAGINLNRSRTAPISG 189

Db 143 RQTADROOQLRARSVNGSQGFEDNAILLAQADAEVAVAEAGVABERLNLYADVKAPISG 202

QY 190 FIGOSKVSSEGLTLNAGDTTLATIRQTNPMYVNTQSAEVMKLRRQIAESKLLAADGVI 249

Db 203 VIGRARITEGALVATSGSENLATIQQLDPIYADFQPADILRLRKALQDQGLMTGQNEA 262

QY 250 AVGKPEDDGTVPYPERKGLFPADPVNVESTGOITLRAVPNONTIMLGLYRVLMDQYAV 309

Db 263 EVNLFPDGSRYRPSVGRLLPSEAAVDETTGQVTLRGEPFNPNGDLLPQMYRVQIQQGIQ 322

QY 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPREVTVAQOQGTWMTVTSGLKDGDKYVE 367

Db 323 KAAFAVVPQQAQVORAGGASVLYVNAEDTVEQRVSVRSIGDKWVISEGLDDGDRYVAE 382

QY 368 GISIAGITGAKKVTPEKEMASSENQAAPQSGVQTASEAKTASE 410

Db 383 GFOKTA-PGA-KVKPEPMSQEPDVAAA--AGSEGAAPSETSSSE 421

RESULT 16

RND multidrug efflux membrane fusion protein MexC precursor PA4599 [imported] - Pseudomo
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: AB3072
R/Stover, C.K.; Pham, X.Q.; Errin, A.L.; Micoicuchi, S.D.; Warriner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: AB3072
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-387 <STO>
A/Cross-references: GB:AE004873; GB:AE004091; NID:g9950839; PIDN:AG07987.1; GSPDB:GN001
C/Experimental source: strain PA01
C/Genetics:
A/Gene: mexC; PA4599

Query Match 34.6%; Score 699.5; DB 2; Length 387;
Best Local Similarity 41.5%; Pred. No. 1.9e-35;
Matches 162; Conservative 82; Mismatches 131; Indels 15; Gaps 9;

QY 10 AALAAALVALVSSCGKGDAAQGGQAGREAPAPVGVVTVHPQTVALTELPGRLS 69

Db 9 RIGALMAAIA--LAGCC----PAEEROEAA-EWLP-VEVLTVOABPLAASSELPGRIEP 60

QY 70 LRTADVPAQVGGITQKRLFOEGSVYRAGQPLXOIDSSTYEANLESARQAOLATQAATLAKA 129

Db 61 VRLAEVRPRVSGIIIVRVEGSGLVKEGDLVLRIDRAPFOVRVDSAGETLLRRAQAOLQA 120

QY 130 DADLARYKPLVAEAASVROEYDAAVTAKRSAAAGVKAQAQAIAKSAGINLNRSRTAPISG 189

Db 121 QARVRRREPVLKIQAVSQGFEDNAILLAQADAEVAVAEAGVABERLNLYADVKAPISG 180

QY 190 FIGOSKVSSEGLTLNAGDTTLATIRQTNPMYVNTQSAEVMKLRRQIAESKLLAADGVI 249

Db 181 RIGALVTBGLVGGQCATLMARIQQLDPIYADFQPADILRLRKALQDQGLMTGQNEA 239

QY 250 AVGKPEDDGTVPYPERKGLFPADPVNVESTGOITLRAVPNONTIMLGLYRVLMDQYAV 309

Db 240 ALTLRV-EGTPYERQGLQFADVAVDGCTQIALRGKFAVPDGVLLPQMYRVQIQQGIQ 298

QY 310 DNAFVVPQQAQVTR--GAKDTVMIVNAQGMPEPREVTVAQOQGTWMTVTSGLKDGDKYVE 367

Db 299 NQAILVPRQAVVHSSDSQAVVMVVGADERAESVGVGVNGQSRWQITBGLPEPDRYVIG 358

QY 368 GISIAGITGAKKVTPEKEMASSENQAAPQSGVQTASEAKTASE 410

Db 359 G--LAAVQPGVKTVPKPDGAQAQAQSPAPQ 386

RESULT 17

probable efflux pump Z2509 [imported] - Escherichia coli (strain 0157:H7, substrain EDL9:
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: B85757
R/Berna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
H.; L.; Grobeck, E.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: AB5480; MUID:21074935; PMID:11206551
A/Accession: B85757
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-373 <STO>
A/Cross-references: GB:AE005174; NID:g12515494; PIDN:AA656521.1; GSPDB:GN00145; UMGp:Z250

A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2509

Query Match 34.5%; Score 697.5; DB 2; Length 373;
Best Local Similarity 42.7%; Pred. No. 2.3e-35;
Matches 162; Conservative 66; Mismatches 134; Indels 17; Gaps 7;

Qy 9 MRAAALAAVALVSSCGKGDAAAGGCPAGREAPAPVGVTVHPQTALVELPGLE 68
Db 1 MKYIATSVVAMLLISGC-----DNTOSNNSSPSETF---VGVTTVKSQGVSVSELTGRTS 53
Qy 69 SLRTADVAAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSSTEYANLESAPQATATQATLAK 128
Db 54 AALSBEVAPQVGGIIQKRLFKEGDLVKAGQPLVQIDASSYQAAWNEAABAALQQAQ-A-LVK 112
Qy 129 ADADLA-RYKPLVAAEAVSRQETDAAVTAKSAEAGVAAQAQAIKSAGINLRKSITAPI 187
Db 113 ADCOKAQRYARLVKENGVSQODADADQOSTCAODKASVAKKAALETARINLDMVTVTAPI 172
Qy 188 SGFTGQSVSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEVMKLRRQIAEGKLLAADG 247
Db 173 SGRIGISVTPGALVTASQDTALTTRIGLDIMTVDRSSVDLRLRKQ----SLATNSD 228
Qy 248 VIAVGIFDGTVPYPEKGRLLFADPVNESTGQITLRAVPNDONILMPGLYVRLMDQV 307
Db 229 TMSVSLIEDGTITSEKGRLELEVADESGVTLRAIFPNPQQQLPGMFVRAVDEG 288
Qy 308 AVDNAFVVPQQAIVRGAKD--TWMIYNAOGMEPREVTVAAQQGNTNIVTSGLKDQKRV 365
Db 289 VMEADIIAPQOGVTRDAKGNATLVNKNKREORTLETGETYGDKMLVINGLHSGDRLI 348
Qy 366 VEGISAGITGAKKVTPE 384
Db 349 VEG--SAKVTSGQTVKAVE 365

RESULT 18

G90861
probable efflux pump Ecs1863 [similarity] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
C:Accession: G90861
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc A:Reference number: A59629; PMID:21156231; PMID:11258796
A:Accession: G90861
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <HAV>
A:Cross-references: GB:BA000007; PIDN:BA35286.1; PID:g13361328; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: Ecs1863

Query Match 34.5%; Score 697.5; DB 2; Length 373;
Best Local Similarity 42.7%; Pred. No. 2.3e-35;
Matches 162; Conservative 66; Mismatches 134; Indels 17; Gaps 7;

Qy 9 MRAAALAAVALVSSCGKGDAAAGGCPAGREAPAPVGVTVHPQTALVELPGLE 68
Db 1 MKYIATSVVAMLLISGC-----DNTOSNNSSPSETF---VGVTTVKSQGVSVSELTGRTS 53
Qy 69 SLRTADVAAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSSTEYANLESAPQATATQATLAK 128
Db 54 AALSBEVAPQVGGIIQKRLFKEGDLVKAGQPLVQIDASSYQAAWNEAABAALQQAQ-A-LVK 112
Qy 129 ADADLA-RYKPLVAAEAVSRQETDAAVTAKSAEAGVAAQAQAIKSAGINLRKSITAPI 187
Db 113 ADCOKAQRYARLVKENGVSQODADADQOSTCAODKASVAKKAALETARINLDMVTVTAPI 172

Qy 188 SGFTGQSVSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEVMKLRRQIAEGKLLAADG 247
Db 173 SGRIGISVTPGALVTASQDTALTTRIGLDIMTVDRSSVDLRLRKQ----SLATNSD 228
Qy 248 VIAVGIFDGTVPYPEKGRLLFADPVNESTGQITLRAVPNDONILMPGLYVRLMDQV 307
Db 229 TMSVSLIEDGTITSEKGRLELEVADESGVTLRAIFPNPQQQLPGMFVRAVDEG 288
Qy 308 AVDNAFVVPQQAIVRGAKD--TWMIYNAOGMEPREVTVAAQQGNTNIVTSGLKDQKRV 365
Db 289 VMEADIIAPQOGVTRDAKGNATLVNKNKREORTLETGETYGDKMLVINGLHSGDRLI 348
Qy 366 VEGISAGITGAKKVTPE 384
Db 349 VEG--SAKVTSGQTVKAVE 365

RESULT 19

B82600
precursor of drug resistance protein XF2093 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82600
R:anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82600
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <SIM>
A:Cross-references: GB:AE004025; GB:AE003849; NID:G9107217; PIDN:AAF84892.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.U.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Agencio, M.; Alvarenga, R.; A Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrer, D.M.; Carrel, H as-Neto, E.; Docena, C.; El-Dorri, H.; Facincant, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemp, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2093

Query Match 33.2%; Score 670; DB 2; Length 408;
Best Local Similarity 36.8%; Pred. No. 1.2e-33;
Matches 146; Conservative 90; Mismatches 149; Indels 12; Gaps 5;

Qy 9 MRAAALAAVALVSSCGKGDAAAGGCPAGREAPAPVGVTVHPQTALVELPGLE 68
Db 1 LRVGLAVTVLALATACK-----SEQP---QMPTEVSVLEAKPQTPPIQLDLVRLS 67
Qy 69 SLRTADVAAQVGGIIQKRLFOEGSVYRAGOPLYQIDSSSTEYANLESAPQATATQATLAK 128
Db 68 AYRSADYARVAGIIQKRIYEGTEVEKQPLFOIDAPAPQATLLEAGRIAAAEATYKN 127
Qy 129 ADADLA-RYKPLVAAEAVSRQETDAAVTAKSAEAGVAAQAQAIKSAGINLRKSITAPI 188
Db 128 AKIVADRRRLSPQVYSRSDVDNAEAEKRAAASVQDARAAQONARINLNAVVTAPIS 187
Qy 189 GFIGQSVSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEVMKLRRQIAEGKLLAADG 247
Db 188 GLAQOGVTEBALVSGSSSTLLITVDQIDPLVYVFNMSNDELMDLRQAQHGSGVQLSSDN 247
Qy 248 VIAVGIFDGTVPYPEKGRLLFADPVNESTGQITLRAVPNDONILMPGLYVRLMDQV 307

Db 248 TSTIDVLLSDGSKYPHOGILPFGCATVDPSTGTSLRAVLPRNTHLLIPGAFVTFKANLG 307
 Qy 308 AVDNAFVVPQQAATRGAKDT-VMIYVNAOGMEPREVTAAOOGTNIWTVTSGLKDGKVVV 366
 Db 308 QRRNTLLPGAGVQGDANASAVYLVDKQKVIKRVKVTITGNKDNQMLTAGLSNDRIV 367
 Qy 367 EGI-SIAGITGAKKVTTPKEMASSENQAAAPOSGVOTA 402
 Db 368 DGLQKVEKGAFAKAI PMKPDLISSSAHAAPKPTHTA 404

RESULT 20

S47733
 yHtu protein precursor - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 01-Mar-2002
 C:Accession: S47733; D65149
 R:Plunkett, G.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S47666
 A:Accession: S47733
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <PLU>
 A:Cross-references: EMBL:U00039; NID:G466582; PIDN:AB18489.1; PID:G466650
 A:Experimental source: strain K-12, substrain MG1655
 R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D65149
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-385 <BLAT>
 A:Cross-references: GB:AE00427; GB:U00096; NID:G1789919; PIDN:AACT6538.1; PID:G1789929;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetic:
 A:Gene: yHtu
 C:Superfamily: lipoyl/biotin-binding homology
 C:Keywords: lipoprotein
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-385/Product: yHtu protein #status predicted <MNT>
 F:57-100,172-201/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 32.5%; Score 656; DB 2; Length 385;
 Best Local Similarity 38.5%; Pred. No. 8.2e-33;
 Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

Qy 25 CGKGGDAAGGQGPAGREAPVAVGVTVHPQTVALTVELPGRLESLRTADVRAQVGGIIQ 84
 Db 14 CGAMLTACDDKSAENAMTPEVGVVTLSPGSVNVLSLPGRTVPEVAELRPQVGII 73
 Qy 85 KRLFOESSYVRAQGPLYQIDSSSTEYANLESARQALTAQATLAKADLARYKPLVAEA 144
 Db 74 KRNFIEDKYNQGDSTLYQIDPAPLQAEINSAKSLAALSTASNARITTFNRQASLTKNY 133
 Qy 145 VSRQEVDAATARKSAAGVKAQAIAKSGININRSRITAPISGFIGOSKVSSEGTLLNA 204
 Db 134 VSRQDYATATQULNEAANVTAKAAVEQATINIQYANVTSPITGVSGKSVTVGALVTA 193
 Qy 205 GDTTVALTIRQTNPMYVNTQSAEWMKLRQIAEGKLLAADGVIAVGIRFDDGVYPEK 264
 Db 194 NQADSLVTQRLDPIYDLTQSDVFLRMEKRVASGQIKQVQSTPVQULNENKRSQT 253
 Qy 265 GRLLFADPVVNESTGQITTLRAAVNDONILMPGLYRVLMDOVAVDNAFVVPQQAATRG 324
 Db 254 GTLKFSDPYVDETTGVTLLRAIFPNPGDILLPGMYTALVDEGSRQVLLVPQGVTHNA 313
 Qy 325 --KDTWIVVNAOGMEPREVTVAOOGGTNMIWTSGLKDGKRVVEGIS--IAGITGAKV 380
 Db 314 QGKATALLDKDQVVKLREIASKAIQDQWVTSGLQAGDVRVIVSGLQIRPGIKARAIS 373

Qy 381 TPKEWASSENQ 391
 Db 374 SSGENASTESK 384

RESULT 21

A91178
 probable membrane protein [imported] - Escherichia coli (strain O157:H7, substrain R1MD
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: A91178
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: A91178
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA37816.1; PID:G13363867; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 050952
 C:Genetic:
 A:Gene: EC84393

Query Match 32.4%; Score 655; DB 2; Length 385;
 Best Local Similarity 38.5%; Pred. No. 9.4e-33;
 Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

Qy 25 CGKGGDAAGGQGPAGREAPVAVGVTVHPQTVALTVELPGRLESLRTADVRAQVGGIIQ 84
 Db 14 CGAMLTACDDKSAENAMTPEVGVVTLSPGSVNVLSLPGRTVPEVAELRPQVGII 73
 Qy 85 KRLFOESSYVRAQGPLYQIDSSSTEYANLESARQALTAQATLAKADLARYKPLVAEA 144
 Db 74 KRNFIEDKYNQGDSTLYQIDPAPLQAEINSAKSLAALSTASNARITTFNRQASLTKNY 133
 Qy 145 VSRQEVDAATARKSAAGVKAQAIAKSGININRSRITAPISGFIGOSKVSSEGTLLNA 204
 Db 134 VSRQDYATATQULNEAANVTAKAAVEQATINIQYANVTSPITGVSGKSVTVGALVTA 193
 Qy 205 GDTTVALTIRQTNPMYVNTQSAEWMKLRQIAEGKLLAADGVIAVGIRFDDGVYPEK 264
 Db 194 NQADSLVTQRLDPIYDLTQSDVFLRMEKRVASGQIKQVQSTPVQULNENKRSQT 253
 Qy 265 GRLLFADPVVNESTGQITTLRAAVNDONILMPGLYRVLMDOVAVDNAFVVPQQAATRG 324
 Db 254 GTLKFSDPYVDETTGVTLLRAIFPNPGDILLPGMYTALVDEGSRQVLLVPQGVTHNA 313
 Qy 325 --KDTWIVVNAOGMEPREVTVAOOGGTNMIWTSGLKDGKRVVEGIS--IAGITGAKV 380
 Db 314 QGKATALLDKDQVVKLREIASKAIQDQWVTSGLQAGDVRVIVSGLQIRPGIKARAIS 373
 Qy 381 TPKEWASSENQ 391
 Db 374 SSGENASTESK 384

RESULT 22

B86024
 probable membrane protein yHtu [imported] - Escherichia coli (strain O157:H7, substrain I
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B86024
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Llim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: B86024
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-385 <STO>

A:Cross-references: GB:AE005174; MID:g1518222; PIDN:AAG58654.1; GSPDB:GN00145; UMG:249
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ynfH

Query Match 32.4%; Score 655; DB 2; Length 385;
Best Local Similarity 38.5%; Pred. No. 9,4e-33;
Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;

25 CGKGGDAAGGCGPAGREAPAPVGVVTHPOTVALTVELPGRLESTADYAVQVGGIIQ 84
Db CGAMLTACDDKSAENTAMTPEVGVVTLSPGSVNLSELPGRTVYEAIRPQVGGII 73
QY KRLFOGGSYVAGCPLYQIDSSTYEANLESARQAATLAKADADLARYKPLVAEA 144
Db KRNFIEGKVNQDGLYQIDPAPLQALNSAKGLAKLSTSNARITFNNQASLTKNY 133
QY VSRQETDAVTAKSAEAGVKAAQAIAKSAGINLRSRTIAPISGFIQSKVSBGTLNA 204
Db VSRQDYDPAFTQNLNEAEANVTAKAAVEQATINLQYANVTSPITGVSGKSVTVGALVTA 193
QY GDTVTLATIRQTNPMYVAVTOSASVMTLRQIAEGKLLAADGVIAVGIKPDGTVPEK 264
Db NOADSLVTQRLDPIYVDTOSVQDFLRKKEVASGQIKQVQSGTFVQNLNENGRYSQT 253
QY GRLLFADPVVESTQITLRAAVPNDQNIIMPGLYVRLMDQVADNAFVVPQCAVTRGA 324
Db GTLKESDPTVDETOSVTLRAFPNPNDLPLGMVYVTLVDEGSQNVLVQDESVTHNA 313
QY --KDTVMIVNAGGMEPREVTVAQOQGTNWTISGLKDGKRVVVGIS--IAGITGAKV 380
Db OGKATATLIDKDDVYKLEIREASKAIGQWVVTSGLDGDRVYVGLQIRIRGIRARAIS 373
QY 381 TPKEWASSENO 391
Db 374 SSQENASTESK 384

RESULT 23

AD0423
multidrug efflux protein [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0423
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-386 <KOR>
A:Cross-references: GB:AL590842; PIDN:CAC92712.1; PID:g15981407; GSPDB:GN00175
C:Genetics:
A:Gene: YP03483

Query Match 31.1%; Score 628.5; DB 2; Length 386;
Best Local Similarity 37.5%; Pred. No. 3.9e-31;
Matches 146; Conservative 79; Mismatches 151; Indels 13; Gaps 5;

QY MRAAALAAVALLVSSCGKGGDAAGGCPAGREAPAPVGVVTHPOTVALTVELPGRLE 68
Db 1 MHSKTLIACTVFTLVACDQ--SSSPSAPPSROE---VGAVTTLKQPVTLSSDLSGRTV 53
QY SLRTADYAVQVGGIIQKRLFOGGSYVAGCPLYQIDSSTYEANLESARQAATLAK 128
Db 54 AAMTSEVAPQVYDGIITKRLFTGSEVTVAGQVLYQIDPASVQAADYATLAKALQNVVSKS 113
QY 129 ADADLARKPLVVAEAVSRQETDAVTAKSAEAGVKAAQAIAKSAGINLRSRTIAPIS 188
Db 114 AKLAKORVYALAKENGVSQODADDAQOTSQOALNAVAKETALLETARINLAVTVRAPIS 173

QY 189 GFIGOSKVBEGTLINAGDTVLATIRQTNPMYVAVTOSASVMTLRQIAEGKLLAADGV 248
Db 174 GRISSTVTPGALVTANQOTTLATIRNLDPYVDTOSASQALLLRQOQAGNPTVANA- 232

QY 249 IAVGIRKDDGTVYPEKGRLLFADPVVESTQITLRAAVPNDQNIIMPGLYVRLMDQVA 308
Db 233 -PVQLTLEDGGSVVAHESQLQTEVAVDPAICAVTLRAKFPNPEHQLLPGMFVRASVTVNGV 291
QY 309 VDMAFVVPQCAVTRGAKD--TVMIVNAGGMEPREVTVAQOQGTNWTISGLKDGKRVV 366
Db 292 NNTIILAPQCGITHDAGNATATLVNQQOQVERREVERTERTIDSYMLISRLGAAGDRLIV 351
QY 367 EGISIAGITGAKKVTPEKMWASSENOAAP 395
Db 352 EGTEKVSV--GDYKVPVEVSTTLVPAVEP 378

RESULT 24

EB33393
RND multidrug efflux membrane fusion protein precursor PA2019 [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: EB33393
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.V.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: EB33393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-396 <STO>
A:Cross-references: GB:AE004628; GB:AE004091; MID:g9948018; PIDN:AAG05407.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2019

Query Match 31.1%; Score 628; DB 2; Length 396;
Best Local Similarity 38.7%; Pred. No. 4.3e-31;
Matches 155; Conservative 65; Mismatches 159; Indels 22; Gaps 5;

QY 14 LAAVALVLSGCGKGGDAAGGCPAGREAPAPV--VGAVVTHPOTVALTVELPGRLESLRT 72
Db 13 LAALVALFLGCEBAAD-----AKTAEAPAVGVIVARPAPIGITSSELPGRLAVRQ 65
QY ADVRAVQVGGIIQKRLFOGGSYVAGCPLYQIDSSTYEANLESARQAATLAKADAD 132
Db 66 AEVBARVAGIYTRRLYEAGQDVRAGTVLFOIDPAPLKAALDISRGALARAASHAAADK 125
QY 133 LARVKPLVAEAVSRQETDAVTAKSAEAGVKAAQAIAKSAGINLRSRTIAPISGFI 192
Db 126 LKRYADILIKORALISREYTEREQTDARQALQIASAKALEQARLRIGVATVTAVIDGAR 185
QY OSKVBEGTLINAGDTVLATIRQTNPMYVAVTOSASVMTLRQIAEGKLL--AADGVYAV 251
Db 186 RALVTEGALVGBDSPTLTVREQIDPIYVNFQAPAGVAAHQRAIRRGQVGVADKQIAV 245
QY 252 GIKEDDGTVYPEKGRLLFADPVVESTQITLRAAVPNDQNIIMPGLYVRLMDQVAVN 311
Db 246 RLVLADGSEYVLAGLELFSDLAVDPGTDITAMRALFPNPRELLPGVGVYQLRGRAVPQ 305
QY 312 AFVVPQCAVTRGACDTVM--IVNAGGMEPREVTVAQOQGTNWTISGLKDGKRVVVGIS 370
Db 306 AITVRDALIRTAQSAVVKVNPGLVEDVEVRADTLQGGDWIISRLKGGEWIVENAA 365
QY 371 IAGITGAKKVTPEKMWASSENOAAPQSGVOTASEAKTASEA 411
Db 366 -----QHAAGSSVQAVVRQPASADAPSPLAASRA 394

RESULT 25

Oy 182 RTAPISGFIQSKVSEGLTINA-GDTVLTATIRQTNPMYVNT---OSASEVMKLAROI 237
Db 172 KVTAPISGFIQSKVSEGLTINADSSSLTSVTQDIPVYVNFSTDEEAMIAKLR--- 228
Oy 238 AEGKILADG-VIANGIRFDGTYVPEKGRLLFADPVVNESTGQITTLRAVPNDONILMP 296
Db 229 AERGATGBADRLKIKILFGGKAYDHGKITDFTSSSLDTEGTIGVRAVENPNHRLIP 288
Oy 297 GLYVRLMDQVAVDNAPVPOAVTRGAK-DTVMVINAOGGMEPREVVAOQGNMIVT 355
Db 289 GQFRAELTIDQVADALIVPKALMQSAQGFVYVNDVNEVPVYGAELKMDLIS 348
Oy 356 SGLKDGKVVVEGSIAGITGAKVTPKEMASSENQAAPSGVGTASEAK 406
Db 349 QGLNSGDRVITEGV-IKAVPG-RPVQPV--VQGVDDKAQAEKGEQADKK 395

RESULT 28

probable drug resistance protein [imported] - Sinorhizobium meliloti (strain 1021) magap
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C/Accession: G95375
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe,
; Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Sutzycki, R.; Wells, D.H.; Yen, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A/Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A/Reference number: A95262; MUID:21396509; PMID:11481432
A/Accession: G95375
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-388 <KOR>
A/Cross-references: GB:AE00469; PIDN:AAK6569.1; PID:G14524048; GSPDB:GN00165
A/Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
C/Genetics:
A:Gene: Smal664
A:Gene: plasmid

Query Match 23.9%; Score 482; DB 2; Length 388;
Best Local Similarity 33.5%; Pred. No. 3.2e-22;
Matches 135; Conservative 81; Mismatches 163; Indels 24; Gaps 12;

Oy 6 FK-AMRAAALAAVA-LVLSGCGKGDAAGGAPAGREAPAPVGVTVVHQTVALTYEL 63
Db 2 PKTAFRSVDFVLGVSGLLCSAGDC--VAQTG-----FVGVGMVQVENSPAHF 50
Oy 64 PGRLESITADVRAQVGIIQKRLFOEGSYVRAQGPLOYIDSTYEANLESARAQLATAQ 123
Db 51 VGRVEALNAVDIARAVEGFLERRLFAEGQVNEKGDLLFRTTYELALEDAQATLVGAQ 110
Oy 124 ATLAKADADLARYPLVAEAVSROEYDAAVTAKRSAGYKAAQAIAKAGINLNSRI 183
Db 111 TNPNAROLQORNAL--SQRTVSQAVIEESHAAARDIAFASVLSAQTRVQAELNIGYTHI 169
Oy 184 TAPISGFIQSKVSEGLTINAGDTVLTATIRQTNPMYVNTQASSEVMKLAROI--EGK 241
Db 170 KAPIDGRIGRAVYSGVLSVP--SSPPLARVVQTDPIRVFVSORTIIDL--RTIAGAGAK 227
Oy 242 LLAADGAVIAGIKFDDGTYVPEKGRLLFADPVVNESTGQITTLRAVPNDONILMPGLYVR 301
Db 228 DELAKG-VALKRLRNSNGEPYQSGKLEFFNGEIVDQITTLRLFLNAQSLMFGQVTV 286
Oy 302 VLMDOVAVDNAPVPOAVTRGAKDT-VMIYNAOGGMEPREVVAOQGNMIVTSGIKD 360

Db 287 VIVEPEREEREPVVPVGSVDEQREGRFVLVDGESSRAVRIRASVOGQNVVEEGLQ 346
Oy 361 GDKVVEG---ISAGITGAKVTPKEMASSENQAAPSGVQ 400
Db 347 GKLIVBGLQVSPGAVAEQSVASGD-AATDTAPAPRUSQ 388

RESULT 29

RND multidrug efflux membrane fusion protein MexE precursor PA2493 [imported] - Pseudomonas
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: E83335
R:Stover, C.K.; Pham, X.Q.; Eyrin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: E83335
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-414 <STO>
A/Cross-references: GB:AE004676; GB:AE004091; NID:G9948532; PIDN:AAQ05881.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A:Gene: mexE; PA2493

Query Match 22.8%; Score 460; DB 2; Length 414;
Best Local Similarity 33.2%; Pred. No. 7.7e-21;
Matches 135; Conservative 74; Mismatches 176; Indels 22; Gaps 9;

Oy 14 LAAVALVLTSSCGKGDAAGGAPAGREAPAPVGVTVVHQTVALTYELPGRLESIRTA 73
Db 13 LALAVALVLTSSCGKAPETTG-----MAAPKSVAVAEIQPLNENDEFTGRLAESEV 65
Oy 74 DVRAQVGIIQKRLFOEGSYVRAQGPLOYIDSTYEANLESARAQLATAQATLAKADADL 133
Db 66 ELRRVSGYIDRVAFHFGALVKKGDLLFQIDPRFEAEVRLAQQLARAAQARSVNEA 125
Oy 134 ARYPLVAEAVSROEYDAAVTAKRSAGYKAAQAIAKAGINLNSRIITAPISGFIQ 193
Db 126 QRGERRLASNAISAELELDARTTAAQAEKAAVAAQTQOLDARLNLSTRTAPIDGRVSR 185
Oy 194 SKVSEGTLLNAGDTVLTATIRQTNPMYVNTQASSEVMK---LRQIAEGTLAADGVIA 250
Db 186 AEVTAAGLVNSGE-TLTLTIVSTDKVAYFPDADERVLTLYELARQ--AGDTRSESPPV 242
Oy 251 VGKFDGTYVPEKGRLLFADPVVNESTGQITTLRAVPNDONILMPGLYVRVLMDOVAVD 310
Db 243 LGLSESDGN--PHLGRDLFDQNVNPRGTIRGRAVFDNAKGEFTPLGLYVRLKLVGSKTY 300
Oy 311 NAFVPOQAV-TGAKRTVMIYNAOGGMEPREVVAOQGNMIVTSGIKDGDGVVYEGI 369
Db 301 AATLIDKEAVGTDLGKFFVLVDGNTVYRYTEMGRLEGRLVRSGLSKGRDIVVNGL 360
Oy 370 SIAGITGAKVTPK--EMASSENQA--AAPSGVGTASEAKTSEAE 412
Db 361 Q--RVPRGMQVDPQKVMASADTLATLARKLQSVGSEPPKVAASKD 405

RESULT 30

probable acriflavin resistance protein [imported] - Sinorhizobium meliloti (strain 1021)
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C/Accession: G96007
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9869-9894, 2001
A/Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: G96007
A/Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-373 <KUR>
 A:Cross-References: GB:AL591985; PIDN:CA949727.1; PID:G15141214; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubier, P.; Chaint, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, M.; Vandenbol, M.; Vorholter, F.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104.
 A:Contents: annotation
 C:Genetic: acc;
 A:Gene: acc; SMB21497
 A:Genome: plasmid

Query Match 22.8%; Score 459.5; DB 2; Length 373;
 Best Local Similarity 30.8%; Pred. No. 7.2e-21;
 Matches 118; Conservative 75; Mismatches 157; Indels 33; Gaps 7;

9 MRAAALAAVALVSSCG-----KGGDAAGGPGAPGAPVAVGVVTHPTVALT 60
 1 MRSRVLAAL-----SCGLTTTIGVAGNAYAOEPPAVNVAFAIM-----DRES 48
 61 VELPGRLESLTADVRAOVGGIIQKRLFOEGSVYRAGOLYQIDSTYEANLESARAOA 120
 49 VDLGKVVAVGVKIDIRAVSGFLEKVNFDGOKVSAGTVLVQVEGAGRAALQEI DSG 108
 121 TQATLAKADADLARYKPLVAEAVSROEYDAVTAKSASAGVKAQAIAKSAGININR 180
 109 AAEQORDLAVERBRAQRLIATNTVAQTLIDTANAOVKKAADILRLKSGKONMELNLSY 168
 181 SRTAPISGFIOGSKVSGTLNAGDTVLATIRQTNPMYVNTQASAEVWKLARQIAEG 240
 169 TKILAPFGVGLTVDVGVALV-APDSSLVTLRLDITVEFPVATSLYSYERKEG 227
 241 KLLAADGVIAVGIFDDTVPEKGRLLFADPVNVESTGOITLRAAVNDONIMPGIYV 300
 228 EMSGAN---VSITLPNTDTPYKGTIDFVASTVSGQIDTVTAEFNPGTILLDGLV 284
 301 RVLMDVAVNADVAVPQOAVTRGAKDT-VMIVNAQGMPEPVTVAQOQGNMTWITSLGK 359
 285 RVVLEQSPDQVLAIVPQAVORDQGAFAVMVVDANSKVELRVDVSRSSGQAVVAGL 344
 360 DGDVVVVGISIAIGTAKKVT 382
 345 EGENVITEGVG-----KVRP 359

Db

RESULT 31
 T30829.
 hypothetical protein mexE - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30829
 R:Kohler, T.; Michie-Hameghour, M.; Henze, U.; Gotoh, N.; Curry, L.K.; Pechere, J.C.
 Mol. Microbiol. 23, 345-354, 1997
 A:Title: Characterization of MexE-MexF-Opn, a novel positively regulated antibiotic eff
 A:Reference number: Z20891; MUID:97197179; PMID:9044268
 A:Accession: T30829
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-414 <KOH>
 A:Cross-References: EMBL:X99514; NID:G1707643; PIDN:CAA67866.1; PID:G1707644
 C:Genetic: acc;
 A:Note: mexE

Query Match 22.1%; Score 446; DB 2; Length 414;
 Best Local Similarity 32.7%; Pred. No. 5.5e-20;
 Matches 133; Conservative 73; Mismatches 179; Indels 22; Gaps 9;

14 LAAAVLVLSGCGGDAAGGPGAPGAPVAVGVVTHPTVALTVELPGRLESLRTA 73

Db

13 LAAAVLVLSAGGAPETTG-----MAAPKVSVAEVIQPLNEMWDEFTGRLEAPSV 65
 74 DVRAQVGGIIQKRLFOEGSVYRAGOLYQIDSTYEANLESARAOAATLAKADADL 133
 66 ELRRVSGYIDRVAFEGALVKKDLDLFQIDPRFEAEVRLKLEQOQARAAQARSVNEA 125
 134 ARYKPLVAEAVSROEYDAVTAKSASAGVKAQAIAKSAGININRSRTAPISGFIGQ 193
 126 QGRGLRASNIAISALDARTTAQAKAAVAAVATQOQDARLNLNLSFRITAPIDGRVSR 185
 194 SKVSEGTILNAGDTVLATIRQTNPMYVNTQASAEVWKLARQIAEGKILAADGVIA 250
 186 AEVTAGNLVNSGECTSPK.Y-STDKVAAYFPDADERVFLKVELARQ--AGDTRSESPPY 242
 251 VGIKFDGTVYPEKGRLLFADPVNVESTGOITLRAAVNDONIMPGIYVAVMDQVAVD 310
 243 LGLSSEIDGN--PHLGRDLFDLNDQVNPRTGIRGAHVPDNAGKERTPGILYRLKLVGSKTY 300
 311 NAFVPOQAV-TRGAKDTVMIVNAQGMPEPVTVAQOQGNMTWITSLGKPDGVVREGI 369
 301 AATILIKQAVGTDLGKFFVLVDGDNKTVYRTVEMGPKLBEGLRIVRSGLSKGDRIVNGL 360
 370 SIAGITGAKKVTPEK--EMASSENOA--AAPQSGVQTASAEKTAASEAE 412
 361 Q--RVPRGMQVDPQKVMASADTLATLARLQSGVGDSEPPKVAASKD 405

Db

RESULT 32
 AG2881
 HlyD family secretion protein (imported) - Agrobacterium tumefaciens (strain C58, Dupont
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AG2881
 R:Wood, D.W.; Seubald, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AG2881
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-396 <KUR>
 A:Cross-References: GB:AE008688; PIDN:AL43469.1; PID:G17740974; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetic: acc;
 A:Gene: mexE
 A:Map position: circular chromosome

Query Match 21.9%; Score 442.5; DB 2; Length 396;
 Best Local Similarity 30.8%; Pred. No. 8.4e-20;
 Matches 123; Conservative 81; Mismatches 173; Indels 23; Gaps 9;

10 RAALAAVALVSSCG-----KGGDAAGGPGAPGAPVAVGVVTHPTVALT 61
 7 RRLATGGIGIAGLSVAAGALFDLPITSRNATASTPA--ETPALPVVAKESADVWRME 64
 62 ELPGRLESLTADVRAOVGGIIQKRLFOEGSVYRAGOLYQIDSTYEANLESARAOA 121
 65 EFGRLVAVRVQIRSVNAQIKVHREBALYKESGPTLIDAPVQAAVAGAGGVAS 124
 122 AATLAKADADLARYKPLVAEAVSROEYDAVTAKSASAGVKAQAIAKSAGININRS 181
 125 ABAKVSIAKTELDRGRLLSDNRITISQGLDQROSSFADAAEQALFAAARAAATLTAOLDGYT 184
 182 RITAPISGFIOGSKVSGTLNAGDTV-LVATIRQTNPMYVNTQASAEVWKLARQI--A 238
 185 EITAPVSGRGRILEITAGNLVAGSTSPALTLTIVSNPITASFNASGVAKALAEIPKT 244
 239 EGKLLAADGV-IAVGIFDDTVPEKGRLLFADPVNVESTGOITLRAAVNDONIMPG 297

Db 245 DGAALPALEQIIVEIGTSLDEGR--PIKGTLMHIDNOVASAGTIGVRAIPFNPDRLLPG 302
 Qy 298 LTVRLMDQVAVNDAAFFVPOQAV--TRGAKDPTVMYVNAAGMEPREVTVAAGQGTWIVTS 356
 Db 303 QFVRKGEKPEPRKENVISDRPIGTDQKRFVVDAAENKVSRYRKIKGAPADGCRITDS 362
 Qy 357 GLKDGKVVVVGISITAGITGAKKVTPEKMASSENQAAPO 396
 Db 363 GLAAGDTIVVNGLO--RIRPGATIAPO---AEDKVAASQ 396

RESULT 33

hypochemical protein AGR_C_4505 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: G97657
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: G97657
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-396 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK8216.1; PID:gl5157668; GSPDB:GN00169
 A:Gene: AGR_C_4505
 A:Map position: circular chromosome

Query Match 21.4%; Score 442.5; DB 2; Length 396;
 Best Local Similarity 30.8%; Pred. No. 8.4e-20;
 Matches 123; Conservative 81; Mismatches 173; Indels 23; Gaps 9;

Qy 10 RAAALAAVALVLSGC-----KGGDAAGGQGPAGREAPVAVVTPVTPVATLV 61
 Db 7 RRAALTGAGIGLAMSVAAGALFFDLPTSRNATVASTPA--ETPAIPVTVAKVESRDVME 64
 Qy 62 ELPGRLBRLTADVAQVGGIIOKRLFOEGSVYRAGOLYQIDSTYEANLESARAOAT 121
 Db 65 EFSGRLAVDVQIRSRVAGQIKAVHFREGALVKEGDLFTIDPAPYQAAVAGEGVVAS 124
 Qy 122 AATLAKADADLARYKPLVAEAVSROEYDAVATKSAEAGVKAQAIAIKSAGINLRS 181
 Db 125 AEAKVSLKTELDKRRISDNKRTISQSLDQROSFADAEQOLRRARALTTAQLDGYT 184
 Qy 182 RITAPISGFIGOSKVSSEGLTLNAGDTT-VLATIRQTNPMYVNVTSASEVWKLRLQI--A 238
 Db 185 EIIAPVSGVRGRIETIAGNLVAAGSTSPALTTIVSVNIVASFNASEGVAKALAEIPKT 244
 Qy 239 EGKLIADGV-IANGIKRDKDGTVPKGRLLPADPVNESGQITLRAVNDQILMPG 297
 Db 245 DGAALPALEQIIVEIGTSLDEGR--PIKGTLMHIDNOVASAGTIGVRAIPFNPDRLLPG 302
 Qy 298 LTVRLMDQVAVNDAAFFVPOQAV--TRGAKDPTVMYVNAAGMEPREVTVAAGQGTWIVTS 356
 Db 303 QFVRKGEKPEPRKENVISDRPIGTDQKRFVVDAAENKVSRYRKIKGAPADGCRITDS 362
 Qy 357 GLKDGKVVVVGISITAGITGAKKVTPEKMASSENQAAPO 396
 Db 363 GLAAGDTIVVNGLO--RIRPGATIAPO---AEDKVAASQ 396

RESULT 34

hypochemical protein b2243 [imported] - Escherichia coli (strain O157:H7, substrain
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B85834
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grochbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B85834
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-464 <STO>
 A:Cross-references: GB:AE005174; NID:912516280; PIDN:AAG57137.1; GSPDB:GN00145; UWGP:232
 A:Experimental source: strain O157:H7, substrain EDL933
 A:Gene: Z3243

Query Match 21.4%; Score 432; DB 2; Length 464;
 Best Local Similarity 31.9%; Pred. No. 4.5e-19;
 Matches 118; Conservative 63; Mismatches 169; Indels 20; Gaps 6;

Qy 34 GGGPAGREAPVAVVGVTVTPVTPVATLVLPGRLESRTADVAQVGGIIOKRLFOEGSY 93
 Db 98 GGRGMSGSLAPVQAATVAVQAVPRYLTGLITTAANTVTVRSRVGQMLAHFQEGXQ 157
 Qy 94 VRAGPIYQIDSTYEANLESARAOATDAQTLAKADADLARYKPLVAEAVSROEYDAA 153
 Db 158 VKAGDLAEIDPSFKVLAQAQGLAKDKKATLTNARDLARYQOLAKTNLVSQEBDAQ 217
 Qy 154 VTAKSAEAGVKAQAIAIKSAGINLNSRITAPISGFIGOSKVSSEGLTLNAGDTTVLATI 213
 Db 218 QALVSEGTGKADAEAVASQQLDWSRITAPVDGRVGLKQVDVNGQISGDTTGIIVI 277
 Qy 214 RQTNPMYVNVTSASEVWKLRLQIABEGLLAADGVAVGI-KFPGDTVPKGRLLPADP 272
 Db 278 TQTHIDLLFTLPSDIATVQAQKAGPLVEMDRNRSKSEGT-----LTLSDN 330
 Qy 273 VVNESTGOITLRAVPNDONILMPGLVY--RVLMDOVAVNDAAFFVPOQAVTRGAK-DTVM 329
 Db 331 QIDTITGKIKAFPNQDPAFPNQVFNARMLVD--TEQNAVIVPFAALOMGEGHFVW 388
 Qy 330 IVNAQGMEPREVTVAAGQGTWIVTSGLKDGKRVVEGI-----SIAGITGAKKVT 382
 Db 389 VLNSGNKVSXKLTVPGIQDSQKVIRAGISAGDRVTVGDRLTEGAKVEVEAQSATTP 448
 Qy 383 KEWASSENQ 392
 Db 449 EKAATSRREYA 458

RESULT 35

hypochemical protein b2074 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: A64974
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 .A.; Rose, D.V.; Mau, B.; Snao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: A64974
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-464 <BLAT>
 A:Cross-references: GB:AE000297; GB:U00096; NID:q1788382; PIDN:AACT5135.1; PID:q1788389;
 A:Experimental source: strain K-12, substrain MG1655

Query Match 21.4%; Score 432; DB 2; Length 464;
 Best Local Similarity 32.2%; Pred. No. 4.5e-19;
 Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

Qy 34 GGGPAGREAPVAVVGVTVTPVTPVATLVLPGRLESRTADVAQVGGIIOKRLFOEGSY 93
 Db 98 GGRGMSGSLAPVQAATVAVQAVPRYLTGLITTAANTVTVRSRVGQMLAHFQEGXQ 157
 Qy 94 VRAGPIYQIDSTYEANLESARAOATDAQTLAKADADLARYKPLVAEAVSROEYDAA 153

Db 158 VKAGDLAEIDPSQFKVLAQAQGLADKATLANARDLARYOOLAKTNLVSROELDAQ 217
Qy 154 VTAKRSAAGVYKAAQAIKSGINLNSRRTAPISGFIGOSKVSSEGLLNAGDTTVLATI 213
Db 218 QALVSETEGTTKADKASVASAQLQIDMSRITAPVGRVGLKQVDVGNQISSGDTTGIIVI 277
Qy 214 ROTNPMYVNTQASAEVWKLRQIAEGKLLADGVIAVGI-KFDDGTIVPEKGRLLFPDP 272
Db 278 TQTHPIDLFTLPESDIAITVVOAQAKGPLVEAWDRTNKSLSEGT-----LTLSDN 330
Qy 273 VNNESTQITLRAAVPNDQNTIMPEGLYV--RYLMDQVAVDNAFVVPQOAVTRGAK-DTVM 329
Db 331 QIDATTTIKVKARPNNDODLFPNQFVNAFMLVD--TEQNAVVIPTALQMGNEGHFW 388
Qy 330 IVNAQGMPEPREVTVAAQOGTNTWITSGIKDQKVVVEGI-----SINGITGAKKVT 382
Db 389 VLNSENKVSXHLVTPGIDQSKVIRAGISAGDRVTVDGIRLTEGAKVEVEAQSATTP 448
Qy 383 KEWASSENOA 392
Db 449 EEKATSRREYA 458

RESULT 36

AC0771.
Probable efflux system protein STY2339 [imported] - Salmonella enterica subsp. enterica
C/Species: Salmonella enterica subsp. enterica serovar Typh
A/Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AC0771
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
Th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moulé, S.; O'Gaora, P.
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <P>
A:Cross-references: GB:AL513382; PIDN:CAD02489.1; PID:G16503353; GSPDB:GN00176
C/Genetics:
A:Gene: STY2339

Query Match 21.3%; Score 431; DB 2; Length 413;
Best Local Similarity 31.9%; Pred. No. 4.5e-19;
Matches 123; Conservative 64; Mismatches 170; Indels 28; Gaps 8;

Qy 32 AAGGQAPAGREAPFVGVVTHPQVATLVELPGRLESRTADVRAQVGGIIOKRLFOEG 91
Db 46 AAGRRGRMGROGLAPVQAATATTQAVPRVLSGLGTVAANTVTVRSRVDDQLALAHQEG 105
Qy 92 SYRAGQPIYQIDSSSTEANLESRAQLATQAATLADADLARYKPLVAEAVSROEYD 151
Db 106 QQVNAAGDLAQIDPSQFKVLAQAQGLADKATLANARDLARYOOLAKTNLVSROELD 165
Qy 152 AAVTAKSAAGVYKAAQAIKSGINLNSRRTAPISGFIGOSKVSSEGLLNAGDTTVLA 211
Db 166 AQAALVNETGTTKADKASVASAQLQIDMSRITAPVGRVGLKQVDVGNQISSGDTTGIIV 225
Qy 212 TIRQTNPMYVNTQASAEVWKLRQIAEGKLLADGVIAVGIKFDGTIVP-EKGRLLFA 270
Db 226 VITQTHPIDLFTLPESDIAITVVOAQAKGPLVEA-----WDRTNKSLSEGLVLSL 278
Qy 271 DPVNVSTQITLRAAVPNDQNTIMPEGLYV--RYLMDQVAVDNAFVVPQOAVTRGAK-DT 327
Db 279 DNOIDPTTGTIKVKARPNNDODLFPNQFVNAFMLVD--TEQNAVVIPTALQMGNEGHFW 336
Qy 328 VMIYNAQGMPEPREVTVAAQOGTNTWITSGIKDQKVVVEGISAGITGAKKVTPEKMAS 387
Db 337 VWLINDENNVSKKRVKIGIODNNRVVISAGLSAGDRVVTVDGID-RLTEGAKV----- 387

Qy 388 SENQAAAPQSGVQTASDAKTASENE 412
Db 388 ---EVEPQT---TMADEKSPSRHE 406

RESULT 37

B90989
Probable membrane protein [imported] - Escherichia coli (strain O157:H7, substrain R1MD
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: B90989
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90989
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-464 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836305.1; PID:G13362351; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C/Genetics:
A:Gene: Ec82882

Query Match 21.3%; Score 431; DB 2; Length 464;
Best Local Similarity 31.9%; Pred. No. 5.2e-19;
Matches 118; Conservative 63; Mismatches 169; Indels 20; Gaps 6;

Qy 34 GQAPAGREAPFVGVVTHPQVATLVELPGRLESRTADVRAQVGGIIOKRLFOEGSY 93
Db 98 GGRGRMGROGLAPVQAATATTQAVPRVLSGLGTVAANTVTVRSRVDDQLALAHQEGQ 157
Qy 94 VRAQGPPIYQIDSSSTEANLESRAQLATQAATLADADLARYKPLVAEAVSROEYDAA 153
Db 158 VKAGDLAEIDPSQFKVLAQAQGLADKATLANARDLARYOOLAKTNLVSROELDAQ 217
Qy 154 VTAKRSAAGVYKAAQAIKSGINLNSRRTAPISGFIGOSKVSSEGLLNAGDTTVLATI 213
Db 218 QALVSETEGTTKADKASVASAQLQIDMSRITAPVGRVGLKQVDVGNQISSGDTTGIIVI 277
Qy 214 ROTNPMYVNTQASAEVWKLRQIAEGKLLADGVIAVGI-KFDDGTIVPEKGRLLFPDP 272
Db 278 TQTHPIDLFTLPESDIAITVVOAQAKGPLVEAWDRTNKSLSEGT-----LTLSDN 330
Qy 273 VNNESTQITLRAAVPNDQNTIMPEGLYV--RYLMDQVAVDNAFVVPQOAVTRGAK-DTVM 329
Db 331 QIDATTTIKVKARPNNDODLFPNQFVNAFMLVD--TEQNAVVIPTALQMGNEGHFW 388
Qy 330 IVNAQGMPEPREVTVAAQOGTNTWITSGIKDQKVVVEGI-----SINGITGAKKVT 382
Db 389 VLNSENKVSXHLVTPGIDQSKVIRAGISAGDRVTVDGIRLTEGAKVEVEAQSATTP 448
Qy 383 KEWASSENOA 392
Db 449 EEKATSRREYA 458

RESULT 38

A83330
Probable RND efflux membrane fusion protein precursor PA2528 [imported] - Pseudomonas aer
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83330
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: A83330
A:Status: preliminary
A:Molecule type: DNA

• Tue Sep 9 08:55:33 2003

us-09-889-756a-2.rpr

Page 17

Search completed: September 8, 2003, 14:02:44
Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 14:02:17; Search time 26 Seconds
(without alignments)
2175.735 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019

Sequence: 1 MAFYAFKMAAALAAVAL.....AAPQGVGTASEAKTASEAE 412

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 513375 seqs, 137303645 residues

Total number of hits satisfying chosen parameters: 513375

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	432	21.4	464	9	US-09-912-020-281
2	370.5	18.4	202	11	US-09-975-719-149
3	175	8.7	659	10	US-09-738-626-6815
4	169	8.4	299	9	US-09-815-242-10168
5	169	8.4	309	9	US-09-815-242-13850
6	161.5	8.0	285	9	US-09-815-242-11913
7	156.5	7.8	390	9	US-09-815-242-11146
8	151.5	7.5	399	12	US-09-769-744A-94
9	138.5	6.9	355	9	US-09-741-669-435
10	127.5	6.3	329	11	US-09-988-067B-52
11	123.5	6.1	329	8	US-08-945-038-4
12	123.5	6.1	407	9	US-09-784-208-4
13	122.5	6.1	407	14	US-10-078-107-2
14	122.5	6.1	407	14	US-10-077-751-2
15	122.5	6.1	407	15	US-10-315-023-4

16	122.5	6.1	407	15	US-10-315-023-10	Sequence 10, Appl
17	120.5	6.0	1180	12	US-10-193-764-61	Sequence 61, Appl
18	120.5	6.0	1188	12	US-10-193-764-59	Sequence 59, Appl
19	119.5	5.9	697	15	US-10-156-761-9338	Sequence 9338, Ap
20	118.5	5.9	407	15	US-10-315-023-11	Sequence 11, Appl
21	118	5.8	382	15	US-10-156-761-12596	Sequence 12596, A
22	118	5.8	1741	10	US-09-971-536-68	Sequence 68, Appl
23	116	5.7	607	11	US-09-308-207-59	Sequence 59, Appl
24	115	5.7	340	15	US-10-156-761-13234	Sequence 13234, A
25	114.5	5.7	1208	15	US-10-156-761-13251	Sequence 13251, A
26	114	5.6	837	9	US-09-815-242-5883	Sequence 5883, Ap
27	114	5.6	875	9	US-09-815-242-13080	Sequence 13080, A
28	114	5.6	2434	9	US-09-815-242-5835	Sequence 5835, Ap
29	114	5.6	6281	9	US-09-815-242-12996	Sequence 12996, A
30	112.5	5.6	272	15	US-10-156-761-12370	Sequence 12370, A
31	112.5	5.6	405	9	US-09-815-242-10096	Sequence 10096, A
32	112.5	5.6	405	14	US-10-078-107-6	Sequence 6, Appl
33	112.5	5.6	405	14	US-10-077-751-6	Sequence 6, Appl
34	112	5.5	542	9	US-09-741-669-414	Sequence 414, App
35	112	5.5	578	9	US-09-159-469-50	Sequence 50, Appl
36	112	5.5	578	9	US-09-798-042-585	Sequence 5885, Ap
37	111	5.5	1029	9	US-09-815-242-5885	Sequence 13083, A
38	111	5.5	1048	9	US-09-815-242-13083	Sequence 1, Appl
39	111	5.5	1525	9	US-09-782-714-1	Sequence 7965, Ap
40	111	5.5	7746	15	US-10-156-761-7965	Sequence 16, Appl
41	109	5.4	542	11	US-09-308-207-16	Sequence 416, App
42	109	5.4	596	12	US-10-238-075-416	Sequence 222, App
43	108.5	5.4	338	10	US-09-881-752A-222	Sequence 1994, Ap
44	108.5	5.4	482	12	US-10-017-161-1994	Sequence 14197, A
45	108.5	5.4	534	15	US-10-156-761-14197	

ALIGNMENTS

RESULT 1

US-09-912-020-281

Sequence 281, Application US/09912020

Patent No. US2002004592A1

GENERAL INFORMATION:

APPLICANT: Zyskind, Judith

APPLICANT: Ohlsen, Kari L.

APPLICANT: Tirawick, John

APPLICANT: Forsyth, R. Allyn

APPLICANT: Froelich, Jamie M.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

FILE REFERENCE: ELITRA 001DV1

CURRENT APPLICATION NUMBER: US/09/912,020

CURRENT FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: 09/492,709

PRIOR FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: 60/117,405

PRIOR FILING DATE: 1999-01-27

NUMBER OF SEQ ID NOS: 485

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 281

LENGTH: 464

TYPE: PRT

ORGANISM: E. Coli

US-09-912-020-281

Query Match 21.4%; Score 432; DB 9; Length 464;

Best Local Similarity 33.2%; Pred. No. 2.1e-30;

Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

DB 98 GGRGRMSGRLPAVOATAVEQAVPRYLTLGTTANTVTVRSDVGGDLALAFQSGQQ 157

QY 94 VRAQGPVQIDSSYNEANLESARQAOLATLAADADLARYKPLVAEAVSROEYDAA 153
D 158 VYAGGLLAIBIDPSOFKVALAQAQOLAKDKATLANARRDLARYQOOLAKTNLSROELDAO 217
QY 154 VTAKSAAGVYKAAQAIAKSGININRSRITPISGFIGOSVSGTILNAGDTVLATI 213
D 218 QALVSETGTIVADASVASAOLQDWSRITAPVGRVGLKVDVNGOISSGDTGIYVI 277
QY 214 ROTNMYVNVVQOSASEVWKLRLQAEGLAADGVIAVGI -KFDGDTYPEKGRLLFADP 272
D 278 TQTHIDLVLTPESDIAITVQAQAKPLVVEADNRNKSLSGDT-----LISLDN 330
QY 273 VNNESTGQITLRAAVPNDONIMPELVY--RYLMDQAVDNAFVVPQAVTRGAK-DTVM 329
D 331 QIDATGTIKYARFNODDALFPQOFVNAKVLVD--TEQNAVVIPTAALQWNEGHFVW 388
QY 330 IVNAGGMEPREVYTAQOQGTWVYTSGLKQDQKVVVEGI-----SIAGITGAKKTP 382
D 389 VLNSENKYSKHLVYTGIDQSQKVIIRAGISAGDRVVTGIDRLTEGARVEVEAQSATTP 448
QY 383 KEMASSENQA 392
D 449 EKATSRREYA 458

RESULT 2

US-09-975-719-149
; Sequence 149, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-149

Query Match 18.4%; Score 370.5; DB 11; Length 202;
Best Local Similarity 37.9%; Pred. No. 2,1e-25;
Matches 80; Conservative 43; Mismatches 73; Indels 15; Gaps 5;
QY 205 GDTTVALITROTNPYVNVVQOSASEVWKLRLQAEGLK-LAADGVIAVGIKFDGTYPE 263
D 1 GQANAMATVQQLDIDYVDVTPSTALLMRRELASGLERAGDNAAKSLKLEDSQYPL 60
QY 264 KGRLLFADPVNNESTGQITLRAAVPNDONIMPELVYVMDQAVDNAFVVPQAVTRG 323
D 61 EGRLEFSVSVDREGSVTLRAVFPNPNELLPGFVNAQLOEGVKQKATILAPQOGVTRD 120
QY 324 AKD--TWAIIVNAGGMEPREVYTAQOQGTWVYTSGLKQDQKVVVEGISIA--GITGAKK 379
D 121 LKGGATLAVNAOKMVELRVIKADRVIGDKVLVTEGLNAGDKIITEGIFQVPGVVE-VKT 179
QY 380 VTPKEMASSENQAAAPQSGVQTASEAKTASE 410
D 180 VPAKQVVASQKADAP-----AKTDSK 201

RESULT 3
US-09-738-626-6815
; Sequence 6815, Application US/09738626

; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6815
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6815

Query Match 8.7%; Score 175; DB 10; Length 659;
Best Local Similarity 22.1%; Pred. No. 4.6e-07;
Matches 102; Conservative 81; Mismatches 194; Indels 84; Gaps 16;

QY 5 AFKMRRAALAAVALVSSCGKGDAAQCGPAGRAPVYGVVTHPTVALTYBL- 63
D 223 ALKARQADAAAEIERADFGILNN-----DRSNLDVIGLDERESLASAESELA 274
QY 64 ---PGRLESIRTDVAOVGGIIOKRLFOGGSYVRAGOPVYOIDSSTYNEANLESARAOL 119
D 275 QARAQGLEAVAAAE--AKVAGLEOSTASKTSTPSPDOYTLQ---STRLEAEKRYVA 328
QY 120 ATQA-----TLAKADADLARYKPLVAEAVSROEYDAV--TAKRSAAGVKA 166
D 329 STBALIARIYIDSIGKVDSELAQAQRAV-AEASHAQ-DALGLETQLOSTQHOLEA 386
QY 167 AQAAIKSA-----GINNRKRITAPISGFI-----GQSKVSEGTILN 203
D 387 QSSAIDAAAGLASVDNEAATRSTQLRMDINNTVRSYSGIVSSVQAAGQAPAAAGALLS 446
QY 204 AGDPT---VLATITROTNPYVNV-----TQSASEVWKLRLQAEGLKLAADVIAVGIK 254
D 447 VADSEIKITANVVEAIEISNVTIGSRVTFPTPSTGTKEPRKSVSPILAA----- 497
QY 255 FDDGTVPEKGRLLFADPVNNESTGQITLRAAVPNDONIMPELVYVMDQAVDNAFV 314
D 498 ---AASAPATGEGAAATTTNDVTPEIISVYGDGEGNLGGSARVRIYHEIAPHVLT 554
QY 315 VPQAVTRG--AKDTMIIVNAGGMEPREVYTAQOQGTWVYTSGLKQDQKVVVEGISI 371
D 555 VPLEAVYKNDGKXAVIISIDNKKVEVEVKTAEISDFDAVSGAGISEBARVLTQGNV 614
QY 372 AGITGAKKVTPEKEMASSENQAAAPQSGVQTASEAKTASE 412
D 615 RGLIGE---TVKLDHDTVBOAAAFSPADPPDPAAPVSAK 652

RESULT 4
US-09-815-242-10168
; Sequence 10168, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeiselbeck, Robert


```
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10168
LENGTH: 299
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10168
```

Query Match 8.4%; Score 169; DB 9; Length 299;
Best Local Similarity 25.4%; Pred. No. 5e-07;

Matches 71; Conservative 43; Mismatches 105; Indels 60; Gaps 9;

```
52 VHPQVLAIVELPRLSLRTADYRAQVGGIIQKRLQEGSYVAGQPLYQIDSTYEAN 111
54 IRAQVSTITPOVSGRIYELNKD-----NOLVAGDLITIDKTPPOIA 97
112 LESRAQLATQAATLADADLARYKPLVAAEAVSROEYDAVATKRSAEAGVAAQA 171
98 ELNQAQLAKAQSDLAANNANRRHL-SGNFISAEELDTANLVAMQASVDAAQATL 156
172 KSAQINLNRSKITPPIGSGVSGSTLINAQDTVLATIRQTNPMYVNTQSASEVM 231
157 KQAWQLAQTEIRAPVSGVWNTLTRIGDYADTG-----KPLFALVDSHSFYVI 205
232 -----KLRRQIAGK-----LLAADG-----VIAVGIKPDGTVYPEKGRLLFADPVN 275
206 GYFEETKL-RHIREGAPAQITLYSDNKTLOGHVSIGALYDQVESDSSITL----PDVK 260
276 ESTGQITLRAAVPNDQNLMPGLYRVLMDOVAVDNAFV 314
261 PNVFVWRLAQRF-----VFALDKVPDVTLLV 288
```

```
RESULT 5
US-09-815-242-13850
Sequence 13850, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
```

```
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13850
LENGTH: 309
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(309)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13850
```

Query Match 8.4%; Score 169; DB 9; Length 309;
Best Local Similarity 24.5%; Pred. No. 5.2e-07;
Matches 63; Conservative 45; Mismatches 103; Indels 46; Gaps 7;

```
75 VRAQVGGIIQKRLQEGSYVAGQPLYQIDSTYEANLESRAQLATQAATLAA--DAD 132
70 VTPQVSGSITQNLNKDQFNAGDVLFVINDKTPHIALNQAQLAKAQSDLAANNHAD 129
133 LARYKPLVAAEAVSROEYDAVATKRSAEAGVAAQAIAISAGINLNRSRTAPISGFIG 192
130 RRRH---LSRNYISAEELDSANLVAMQASVDVALLTLOQWQLSQTEVKAFVSGWVT 186
193 QSKVSEGTLLINAGDTVLATIRQTNPMYVNTQSASEVM-----KLRR-----QI 237
187 NLSTRIDVASTG-----KPLFALVDSHSFYVMGYFEETKLRHIREGPAITL 235
238 AEGKLLAADGVIAVGIKPDGTVYPEKGRLLFADPVNNESTGQITLRAAVPNDQNLMPG 297
236 YSGNVKLQGHVGSIGRAIYDQVESDGLV----PDIKPVVFWRLAQRF----- 282
298 LVYRVLMDOVAVDNAFV 314
283 --VRIEFDALPDITLLV 297
```

```
RESULT 6
US-09-815-242-11913
Sequence 11913, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
```

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11913
LENGTH: 285
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11913

Query Match 8.0%; Score 161.5; DB 9; Length 285;
Best Local Similarity 26.0%; Pred. No. 2.2e-06;
Matches 69; Conservative 40; Mismatches 101; Indels 55; Gaps 10;

71 RTADVRA-----OVGGIIORLFOEGSYVYAGQPLVQIDSTYEANLESARQALATAQ 123
35 RDARVRAVVVVVAPVPSGVWTDLEVKNDQVYKGVLRIDORQANLEQARAFAETRH 94
124 ATLAKADADLARYPEL-----VAAEAVSRQEDDAVATKRSAGVKAQAQAIKSNINL 179
95 QOYLRLQNEAARRSLGIGALISAEDEKENAQINAAI-----ARSEYQELAQVKIAELNLIK 149
180 RSRITAPISGRTGQSKVSEGTILNAGDTTVALITIQTN---PMYVNTQASAEVKKLRQ 236
150 RSEILAAANGQVTNLRLEQGNATAGQ-AVVALVDQSGSYVAYEEETKLPGIRVGMBAQ 208
237 IAEGLLAADVGLAIGIFEDGTVPPEKGRLLFADPVNVESTGQITLRAAVPNDQ---NI 293
209 V---RLMSGDQPI-----DGT-----VESISSGITIDRNSITPDGQLIANV 244
294 -----LMPGLYVRLMDQVAVD 310
245 EPTFWVWVRLAQRIPIVRIRLDQVPAD 269

RESULT 7
US-09-815-242-11146
Sequence 11146, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11146
LENGTH: 390
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-11146

Query Match 7.8%; Score 156.5; DB 9; Length 390;
Best Local Similarity 21.6%; Pred. No. 9.8e-06;
Matches 71; Conservative 51; Mismatches 115; Indels 91; Gaps 10;

75 VRAQVGGIIORLFOEGSYVYAGQPLVQIDSTYEANLESARQALATA----- 122
64 VSSQVAGNVAKINDMNDKVHAGDILVELDDTNKAKLSFEQAKSNLAVARQVEQLGFTVQ 123
123 -----QATLAKADADLARYPELVAAEAVSRQEDDAVATKRSAGVKAQA----- 168
124 QLOSAVHANEISLAQAGNLARVQLEKMGALIDKESFQHAKEAVELAKANINASKQGLAA 183
169 -----AAKSGINLRSRITAPISGRTGQSKVSEGTILNAGDTT 208
184 NQALLRNVPLREQOIQNAISLKQAMINLORTIRSIDIVYARRNVQQAASVGCAL 243
209 VLAITRQTNPMYVNTQASAEVKKLRQI-----AEGKLLAADGVIAVGIFEDGTTYPE 263
244 MAVVSNQEMLEANFKETQLTNMRIGQPVKIHFDLYGKNEKFEDEVIN-GIEMGTGNAF-- 300
264 KGRLLFADPVNVESTGQITLRAAVPNDQNIIMPGLYVRLMDQVAVDQVAVNFAFVPPQ----- 318
301 --SL--PSQNAATGMWIKVQVRP-----VRIKLD-----PQGFETTP 334
319 -AVTRGAKDTVMIVNAQGM-----EPR 340
335 LRIGLSATKVRISDSSGAMLRKTERK 362

RESULT 8
US-09-769-744A-94
Sequence 94, Application US/09769744A
Publication No. US20030134407A1
GENERAL INFORMATION:
APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hanniffey, Sean B
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P2112WO
CURRENT APPLICATION NUMBER: US/09/769,744A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/GB99/02452
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: GB 9816336.3
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125329
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 94
LENGTH: 399
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-744A-94

Query Match 7.5%; Score 151.5; DB 12; Length 399;
Best Local Similarity 24.9%; Pred. No. 2.9e-05;
Matches 98; Conservative 64; Mismatches 163; Indels 69; Gaps 18;

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945.038
FILING DATE: 23-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00225
FILING DATE: 19-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN2575
FILING DATE: 21-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN3931
FILING DATE: 03-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN7565
FILING DATE: 16-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 017227/0133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-038-4

Query Match 6.1%; Score 122.5; DB 8; Length 329;
Best Local Similarity 22.8%; Pred. No. 0.0087;

Matches 63; Conservative 37; Mismatches 109; Indels 67; Gaps 6;

QY 63 LPRGLESRTADVRAQVGGIIQKRLFOEGSVYRAGQPIYQIDSSYFEANLESARA-----117
DB 38 LQGFLEA-REYSVSSKVFGRLEKVFVKGDRIKKDDLVFSISSPELEAKQAEGHQA 96
QY 118 -----QATQAATLA-----KADADLAR 135
DB 97 KALSDVVRKGRDETINSABDVQAQAKQATLAKETRYKVDLYONGVASIQKREAVAA 156
QY 136 YKPLVAABAVSRQETDAV-----TAKSAEAGVAAQQAISAGINLRSRTITPISGF 190
DB 157 YESTKYNESAAYQKYMALGGASSESKIAAKAKESAAALGOVNEVESYLKDVATAPIDGE 216
QY 191 IGQSVSSEGLTNAGDTVTYLAITRTQNPVYVNTQSASEVMKLRQIAG-----240
DB 217 VSNVLISGSELSPKGFVPLMTLDKDSWIKTSVPEKYINDFKVGKEP-BGYIPALKRSK 275
QY 241 ---KLLAADGVAVGIRKFDGTVPYPEKGRLLFADPV 273
DB 276 FRVKYLSVWGDFATMKATINNSVTYMKGYEVAIPL 311

RESULT 12
US-09-784-208-4
Sequence 4, Application US/09784208
Patent No. US20010019836A1
GENERAL INFORMATION:
APPLICANT: IZUI, HIROSHI
APPLICANT: ONO, EIJI
APPLICANT: MATSURI, KAZUHIKO
APPLICANT: MORIYA, MIKA
APPLICANT: ITO, HISAO
APPLICANT: HARA, YOSHIIKO
TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR
TITLE OF INVENTION: PRODUCING L-GLUTAMIC ACID
FILE REFERENCE: 0010-0989-0

CURRENT APPLICATION NUMBER: US/09/784.208
CURRENT FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 09/271,438
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: JP 10-69068
PRIOR FILING DATE: 1998-03-18
PRIOR APPLICATION NUMBER: JP 10-297129
PRIOR FILING DATE: 1998-10-19
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 407
TYPE: PRT
ORGANISM: Enterobacter agglomerans
US-09-784-208-4

Query Match 6.1%; Score 122.5; DB 9; Length 407;
Best Local Similarity 20.9%; Pred. No. 0.012;
Matches 83; Conservative 58; Mismatches 154; Indels 103; Gaps 13;

QY 15 AAVALVLSGCKGDAAGGQAPAGREAPAVGVVTHPQTVALTVLPGRLESRTAD 74
DB 14 SVADPTVATWHKKKQGDVSRDE-----VIVEITDKVLEVP-----50
QY 75 VRAQVGGIIQKRLFOEGSVYRAGQPIYQI-----DSTYEANLESARAQATLAQATL 126
DB 51 --ASADGVLEAVLDEGATVTSRQILGRLEKNGSAGKSSAKASNDTPAQRQTASLEE 108
QY 127 AKADA-----DLARYKPLVAABAVSRQETDAVAVTAKRSAAEAVKAAQAIAIK 172
DB 109 ESSDALSPAIRRLIAEHNLDAAQIKGTGVGRLTRDEVEKHLAKPQAEKAAAPAGAT 168
QY 173 SAGINLNRSRITAPISGFIGQSKVSEGLTNAGDTVTYLAITRTQNPVYVNTQSASEVMK 232
DB 169 AQQEVANRSEKRVMTRL--RKVAERLEKKNSTAMLTNNEIN-----MKPIMD 217
QY 233 LRQIAE-----GKLLAADGVAVGIR--FDDGTVPYPEKGRLLFADPVNESTGQITLR 284
DB 218 LRQYGDAPFERKHGVRLL---GFMSFYIKAAVEALKRYPEVNASIDGEDVYVNHVFDVISA 274
QY 285 AAVENDQNIIMPGLYKRVLMQVAVDNAFVVPQAVTRGADVTYITVAQGMREPVTV 344
DB 275 VSTR-----GLVTPLRDVDALSMADI-----EKKIKELAVKVRGDS---KLTV 315
QY 345 AQQGTNWIYVSGKDGDKVVEGISIAGTGAKKTP 382
DB 316 DDLTGNNFTTNG-----GVGSLMSIP 338

RESULT 13
US-10-078-107-2

Sequence 2, Application US/10078107
Publication No. US20020182688A1

GENERAL INFORMATION:

APPLICANT: IZUI, HIROSHI

APPLICANT: HARA, YOSHIIKO

APPLICANT: SATO, MASAKAZU

APPLICANT: AKIYOSHI, NAOKI

TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID

FILE REFERENCE: 219846USO

CURRENT APPLICATION NUMBER: US/10/078.107

CURRENT FILING DATE: 2002-05-13

PRIOR APPLICATION NUMBER: JP 2001-044134

PRIOR FILING DATE: 2001-02-20

NUMBER OF SEQ ID NOS: 8

SOFTWARE: Patentin version 3.1

SEQ ID NO 2

LENGTH: 407

TYPE: PRT

ORGANISM: Enterobacter agglomerans

US-10-078-107-2

Query Match 6.1%; Score 122.5; DB 14; Length 407;

Db 316 DDLTGNFTITNG-----GVFGSLMSTP 338

RESULT 16

US-10-315-023-10

Sequence 10, Application US/10315023

Publication No. US20030119153A1

GENERAL INFORMATION:

APPLICANT: IZUI, Hiroshi

APPLICANT: ONO, Eiji

APPLICANT: MATSUI, Kazuhiko

APPLICANT: MORIYA, Mika

APPLICANT: ITO, Hisao

APPLICANT: HARA, Yoshihiko

TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR PRODUCING L-GLUTAMIC ACID

TITLE OF INVENTION: ACID

FILE REFERENCE: 0010-0989-0

CURRENT APPLICATION NUMBER: US/10/315,023

CURRENT FILING DATE: 2002-12-10

PRIOR APPLICATION NUMBER: US/09/271,438

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: JP10-69068

PRIOR FILING DATE: 1998-03-18

PRIOR APPLICATION NUMBER: JP10-297129

PRIOR FILING DATE: 1998-10-19

NUMBER OF SEQ ID NOS: 15

SOFTWARE: Patentin version 3.1

SEQ ID NO 10

LENGTH: 407

TYPE: PRT

ORGANISM: Enterobacter agglomerans

US-10-315-023-10

Query Match 6.1%; Score 122.5; DB 15; Length 407;

Best Local Similarity 20.9%; Pred. No. 0.012;

Matches 83; Conservative 58; Mismatches 154; Indels 103; Gaps 13;

Db 15 AAVALVSSCGGDAAGGQAGREAPAPVGVVTPVATLVELPGRLESLRAD 74

14 SVADATVATWKKKPPDDAVSRDE-----VYIEIETDKVLEVP----- 50

Qy 75 VRAVGGIITQKRLFOEGSYVRAGQPLYOI-----DSSTYEANLESARQATATQATL 126

Db 51 --ASADGVLEAVLEDEGATVTSRQLTGRLEKNSAGKSSAKAESNDTTPAQRTASLE 108

Qy 127 AKADA-----DLARYPLVAABEVSQOEIDAANTAKRSABAGYKAQAQAIK 172

Db 109 ESSDLSPAIRRLIAEHNLDAQIKGTGVGRLTREDVEKHLANKPQAEKAAAPAGAT 168

Qy 173 SAGIMLNRSRTAPISGFIGSKVSEGLTLNAGDTTVALTIRQTNPMYVNVTSASEYWK 232

Db 169 AQPAPANSSEKRVPTRL--KRVARELLLEAKNSTAMLTTFPEIN-----MKPIMD 217

Qy 223 LRRQIAE-----GKLLADGVIAVGIR--PDDGTVPYEPKGRLLPADPVNVSTGQITLR 284

Db 218 LRKQYGDAPFEKRGVRL--GFMSFYIKAVVEALRKREPVNASIGDEVDVYHNPVDSIA 274

Qy 285 AAVPNDQIIMPLGLYVRVILMDQVADNAFVVPQAVTGAKDTWIVAAOGMEHRETV 344

Db 275 VSTPR-----GLVTPVLRVDVDAISMADI-----EKIKELAVKGRDG---KLTV 315

Qy 345 AQQGQTNMIVTSGLKDGKVVVEGISIAGITGAKKVT 382

Db 316 DDLTGNFTITNG-----GVFGSLMSTP 338

RESULT 17

US-10-193-764-61

Sequence 61, Application US/10193764

Publication No. US20030133943A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yang, Yan-ping

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS

FILE REFERENCE: 1038-1239MIS

CURRENT FILING DATE: 2002-07-12

PRIOR APPLICATION NUMBER: US/10/193,764

PRIOR FILING DATE: 1998-10-07

NUMBER OF SEQ ID NOS: 91

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 61

LENGTH: 1180

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-10-193-764-61

Query Match 6.0%; Score 120.5; DB 12; Length 1180;

Best Local Similarity 20.7%; Pred. No. 0.088;

Matches 89; Conservative 66; Mismatches 178; Indels 97; Gaps 18;

Qy 28 GDAAGGQAGREAPAPVGVVTPVATLVELPGR-----LESLRADVRAQVGII 83

Db 750 GGDVLAVGNISG-----NAVTVTANSGLTI-LAGSTIKGTESITTSQSGNIGKI 800

Qy 84 QKRLFOEGSYVRAGQPL-----YQIDSTYEANLESARQATATQATLAKADADLARYKP 138

Db 801 SKGTVN-----VKATNSLITQADSKIEATEGEANTSKTSIIG--GTSIGTVEVTATEG 853

Qy 139 LVAAEAVSRQETDAVTAKSABEG-----VKAQAALISAGINLNRSR 182

Db 854 LTTQAGSTITGTESVTSQSGNIGMISGKVEVSATKDLITSGSEIKATAGEVNTS 913

Qy 183 ITAPISGFIGSKVSEGLTLNAGDTV-----LATIRQTNPMYVNVTSASEYWK 231

Db 914 ATGITDITSGNTVN--VTANTGDLTVEDAKIDATGCAATLTATSGKL--TTAASSIT 969

Qy 232 KLRQIAEGKLLADGVIAVGIRPD-----GTVEPKGRLLPADP---VVN---EST 278

Db 970 SANNOV--NLSAKDGSIGGINANAVTLNTGALTTVKSSINANSGLVINKADELN 1026

Qy 279 GQITLRAVPNDQIIMPLGLYVRVILMDQVADNAFVVPQAVTGADTWTIVAAOGME 338

Db 1027 GEASGNHTVNVATANGSGSVIATTSRVNI-----TGDLITING--- 1066

Qy 339 PREVTAAQOGTNMIVTSGLKDGKVVVEGI--SIAGITGAKKVPK--EMASSENQAPQ 396

Db 1067 ---LNTISKGINVTLKGVKIDVKYIQPGIASVDEVIEAKRILEKVKDLSDEREALAK 1123

Qy 397 SGVQTASEAK 406

Db 1124 LGVSAVRPAE 1133

RESULT 18

US-10-193-764-59

Sequence 59, Application US/10193764

Publication No. US20030133943A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yang, Yan-ping

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH MOLECULAR WEIGHT PROTEINS

FILE REFERENCE: 1038-1239MIS

CURRENT FILING DATE: 2002-07-12

PRIOR APPLICATION NUMBER: US/10/193,764

PRIOR FILING DATE: 1998-10-07

NUMBER OF SEQ ID NOS: 91

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 59

LENGTH: 1188

TYPE: PRT

ORGANISM: Haemophilus influenzae
US-10-193-764-59

Query Match 6.0%; Score 120.5; DB 12; Length 1188;
Best Local Similarity 20.7%; Pred. No. 0.089;
Matches 89; Conservative 66; Mismatches 178; Indels 97; Gaps 18;

QY 28 GDDAAGGOGPAGREAPAVGVVTVHPQVATLVELPGR-----LESRTADVRAQVGGII 83
DB 758 GGDTLAVANGISG-----NAVTVANGALTT- IAGSTIKGTEISITSSQSGNGIKI 808
QY 84 QKRLFOESSYVRAQPL-----YQDSSTEANLESAPAQLATAQATLAKADADLARKP 138
DB 809 SGKTVN---VKATNSLTTOADSKIEATEGEANVTSKSIIG---GTTSGGTVEVATEG 861
QY 139 LVAAEAVSRQRYDAVATAKRSAG-----VKAQAQAIKSGAGINLNSR 182
DB 862 LTTQAGSITTTGTESTTSSQSGNIGMISGKVEVSATKDLITSSGSIKATAGEVNTS 921
QY 183 ITAPISGFIGOSKVSSEGTLLNAGDTTV-----LATIRQTNPMYVNTQASAEVM 231
DB 922 ATGTIDGTISGNTVN--VTANTGDLTVEDA KIDATGAGATLTLATSGKL--TTKASSSIT 977
QY 232 KLRQIAGSKLLADGVAVGKFPD-----GTVPKGRLLFPDP--VNV---EST 278
DB 978 SANNOV---NLISAKDGSIGGINAMAVTLTGTGALTTVKSSINANSGLTVINAKDAELN 1034
QY 279 GQITLRAVPNDQNLMPGLVVRVLMDOVADNAFVVPQAVTGAKDTWIVAAQSGME 338
DB 1035 GEASGNTVAVANANGSGSVIATTSRVNI-----TGDLITING-- 1074
QY 339 PREYVAQOOGTNNIVTSGLDKDGKVVVEGI-SIAGTGAKKVTPK-EMASSENQAAAPQ 396
DB 1075 ---LNIISKNGINTVLKGVINDVYKIQIOPGLASVDEVIEARKILEKVDLSDEBERALAK 1131
QY 397 SGVOTASPAK 406
DB 1132 LGVSAVRPAE 1141

RESULT 19

US-10-156-761-9338
; Sequence 9338, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIYA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9338
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9338

Query Match 5.9%; Score 119.5; DB 15; Length 696;
Best Local Similarity 22.0%; Pred. No. 0.049;
Matches 97; Conservative 63; Mismatches 162; Indels 119; Gaps 20;
QY 52 VHPQVATLVLELP-----GRLESRTADVRAQV-GGIQKRLFOEGSSYVRAG 97
DB 1132 LGVSAVRPAE 1141

DB 161 LHP--APLVQPLPIGTODGFRGVVDLVMSLMLADGRDTEVSGPPDALARDEAD--RRR 216
QY 98 QPLVQDSSTEANLESAPAQLATAQATLAKADADLAR-----YKPLV 140
DB 217 RQLEEAVALHPLVALEEFCAQATVSAQTLTRALDLRTGTGGVVVLCGSAVRNGIEPLL 276
QY 141 AAEAVSRQRYDAVATAKRSAGVKAQAQAIKSGAGINLNSRITAPISGFIGOSKVSSEGT 200
DB 277 AA-AVALPSPLDPAVRGTEDGAVRERADPAAPFAALVFKVATATGRLTYLVYSGT 335
QY 201 LMANQDTTVLATIRQTNPMYVNTQASAEVMKLRQIABEKLTA----- 244
DB 336 -IGKFTVLDTGARSERIRILRVQDRHTELDRAVA-GDIVAVGPKTARAGSTLCAP 393
QY 245 -----ADGVAVGKFPDGTVPKGRLLFPDPVNVNESTGOITLRAVPNDONI 293
DB 394 AAPLVLEPPTVADPVSAVAEARSST---PTDRLASLALAEEDPSLVRTDPTGTQV 450
QY 294 L--MPGLVVRVLMDOVADNAFVVP-----QQAATRGAKDTV----- 328
DB 451 LSGWGEHLLEVAVEKIRDRGTAVNVRPRVAVRETVARGVSLVYRHKQDGGAGQFAH 510
QY 329 MIVNAQ-----CGMPEYTVAAQOOGTNNI--VTSGLDKDGKVVVEGISIAG--I 374
DB 511 VVLDVEPLESAADGCHGFEFRSAVVGARVQEVYVRAVAGCRD---ALAEV-PLAGHPV 566
QY 375 TGAKKVT-----PKWASSE 389
DB 567 TGLRVVLTGATHPKD---SSS 585

RESULT 20

US-10-315-023-11
; Sequence 11, Application US/10315023
; Publication No. US20030119153A1
; GENERAL INFORMATION:
; APPLICANT: IZUI, Hiroshi
; APPLICANT: ONO, Eiichi
; APPLICANT: MATSUI, Kazuhiko
; APPLICANT: MORIYA, Mika
; APPLICANT: ITO, Hisao
; APPLICANT: HARA, Yoshihiko
; TITLE OF INVENTION: L-GLUTAMIC ACID-PRODUCING BACTERIUM AND METHOD FOR PRODUCING L-GL
; FILE REFERENCE: 0010-0989-0
; CURRENT APPLICATION NUMBER: US/10/315, 023
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US/09/271, 438
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: JP10-69068
; PRIOR FILING DATE: 1998-03-18
; PRIOR APPLICATION NUMBER: JP10-297129
; PRIOR FILING DATE: 1998-10-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-315-023-11

Query Match 5.9%; Score 118.5; DB 15; Length 407;
Best Local Similarity 21.4%; Pred. No. 0.027;
Matches 86; Conservative 55; Mismatches 151; Indels 109; Gaps 15;
QY 15 AAVALVLTSSCGGDAAGGOGPAGREAPAVGVVTVHPQVATLVELPGRLESRTAD 74
DB 14 SVADATVATVHKKRGDAV-----VRDEVLEIETDKVLEVP----- 50
QY 75 VRAQVGGIITKRLFOESSYVRAQPLVQI---DSSTYEALES-----ARAQALTAQ- 123
DB 51 --ASADGILDAVLEDEGTVTTSRQIIGRLREGNSAKGETSAKSEKASTPAORQASLEE 108

124 -----ATLAKADADLARYKPLVAEAVSRQEDAVATAVTAKRSAEAGVKAQAIAIK 172
109 ONNDALSPAIRRLIAEHMIDAAQIKGTGVGSLTREDVEKHLANKPOAEKKAAPAAAGAT 168
173 SAGINLNSRITAPISGIGOSKSEGTULNAGDTTVALTROTPTMTVNTQASSEVMK 232
169 AQPVAANSERKVPMTRL--RRVAERLLEAKNSTAMLTTFEEV--MKPIMD 217
233 LRQIAEGLKLAADGVIAVGIKFDGTV-----YPEGRLLEFADPVVNESTGQITLRAA 286
218 LRQVGEA-FEKRHRIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYVNTQASSEVMK 276
287 VPNDQNLMPGLYFVLMDOVAVDNAFVVPQAVTRGAKD-----TWIYVAGGMEPRE 341
277 TPR-----GLVTFVLRD--VD-----TLGMADIEKKIKELAVKGRD--K 312
342 VTVAGQGTNMTVTSGLKDGKRVVVEGISIAGITAKKVT 382
313 LTVEDLTGNTFTTG-----GVFGSLMSTP 338

RESULT 21

US-10-156-761-12596
; Sequence 12596, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12596
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12596

Query Match 5.8%; Score 118; DB 15; Length 382;

Best Local Similarity 23.4%; Pred. No. 0.028;
Matches 97; Conservative 39; Mismatches 171; Indels 108; Gaps 15;

32 AAGGCG-----PAGR--EAPAPVGVVTHPQTALVEL-----PGRLESRTADYR- 76
6 AAGSGDNRPTELPAIRWDEPPEGPVLVLDOTRLAEFEVELCTDAPALVEAIRTLAARG 65
77 AOVGIIQKRLPQESY-----VRAGPLYOIDSSTVEANLESARAOATQA 124
66 APLUCI-----AGAYVALLAANGFVDEAQAOLSGARPTAVNLAVGRRARAAAYDA 118
125 TLAKADADLARYKPLVAEAVSRQEDAVATAVTAKRSAEAGVKAQAIAIKSAGINLNSRIT 184
119 LAGGDPERRAAALQAARALHREDAEASA--RMAEHGLALDELPGCGHRL----- 169
185 APISGFIQSKVSEGLNAGDTTVALT-----IRCTNPMYVNTQASSEV 230
170 -----LTHCNTALVSGEGTAFAYALAHRGRRLRLMWDTRPLQGARLTAYEA 221
221 MK-----LRQIAEGLKLAADGVIAVGIKFDGTVPEKGRLLFADPVVNESTGQITLR 284
222 ARNGAYTLLTDNAGSLFAGEVDAVLIGAD-----RIADSGSVANKVGSYPL- 270
285 AAVPNDQNLMPGLYFVLMDOVAVDNAFVVPQAVTRGAKDTVMI-VNAGGMEPREVT 343

271 -----AVLARHHPRIIVAPVTTVPDTPDGASIEVBOARAGEVTEVT 314
344 VAQOGTNTMTVTSGLKDGKRVVVEGISIAGITAKKVTPEKMA--SENGAAP 395
315 APQ-----VFVAGAEAGGIPVAPLGTQAVNPAPFVDPBELVATVTEGASVP 363

RESULT 22

US-09-971-536-68
; Sequence 68, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Hayukkala, Ilkka
; APPLICANT: Lubbers, Leonard
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1741
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-68

Query Match 5.8%; Score 118; DB 10; Length 1741;
Best Local Similarity 19.6%; Pred. No. 0.26;
Matches 106; Conservative 75; Mismatches 201; Indels 158; Gaps 22;

22 LSSCGKGDAAGGQAPRGREAP--APVVGVTTHPQTVA-----LYVELPRLSRTAD 74
388 LSDGAKAKLQAATGNNYQLTADDLAKVTGALTITPATTSVDSDVSFEYDKTKASEAAG 447
75 VRAGV-----GGIIQ--KRLQEGSYVRAGOPLYOIDSSTVEANLES----- 115
448 IQATIKLDTGKVVLTDADIIVTNDVNAQGYSTOL--SDGAKAKLQAATGNNYQLTADDL 506
116 -----RAQLATQAATLAKADADLARYKPLVAE--AVSR 147
507 AKVAGTITTPAVTTVDSVDSFEYDKTKASEAKGQATIKLDTGKVVLTDADIIVTN 566
148 OEYDAATYAKSAAAGVKAQAQA-----ISAG--INLNSRITPISG-----F 190
567 DDVNAQVSYQLSDGAKAKLQAATGNNYQLTADDLAKVMTGITTPAAVTAVDLSDFEY 626
191 IGOSKVSSEG-----TLNAGDTTVALTROTPTMTVNTQASA-----SEVMKLRQIAE 239
627 DCKTKASEAKIQAMVNLGETEKTVDLTSADIVANDVNAQVSYQLSDGAKAKLQAAT 686
240 GK--LLAADGVIAVG-----IKFD--DGYVPEKGRLLFADPVVNESTGQ 280
687 GNNYQLTADDLAKVAGTITTPATTSVDSDVSFEYDKTKASEAKGQATIKLGEIEKT 746
281 ITLRBA--VPNDQNLMPGLYFVLMDOVAVDNAF-----VVP 316
747 VDLSSADIIVANDVIV--GKYTSLSDSGSKLQAATGNNYQLTTEVLDRKVSISITTP 804

QY 317 QQAATRG-----AKIVIMLVNAQGMEREPELVVAQOOGTMMIVTSG-----357
 805 AGAATGKDAAFEYDGTAKTSAEKGIALITIDGTEKTVLDAADIYVAEDGVADGKYSY 864
 Db
 QY 358 -LDDGKVVYE-----GISIAGTGAKKATPKEMASSENQAAAPSGVOTASAEAK 406
 865 RLSDGAKSKIOREAGSDHQLTADDAEVNGLTTTPALATADSDNSEVYNGKTKSAEE 924
 Db

RESULT 23

```

1      : Sequence 59, Application US/09308207
2      : Publication No. US2003002232A1
3      : GENERAL INFORMATION-
4      : APPLICANT: MARIA DIAZ-TORRES ET AL.
5      : TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
6      : PRODUCTION OF 1,3 PROPANEDIOL
7      :
8      : NUMBER OF SEQUENCES: 68
9      : CORRESPONDENCE ADDRESS:
10     : ADDRESS: Genencor International, Inc.
11     : STREET: 4 Cambridge Place
12     : 1870 South Winton road
13     : CITY: Rochester
14     : STATE: NY
15     : COUNTRY: U.S.A
16     : ZIP: 14618
17     :
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Diskette
20     : COMPUTER: IBM Compatible
21     : OPERATING SYSTEM: Windows
22     : SOFTWARE: FastSeq for Windows Version 2.0b
23     :
24     : CURRENT APPLICATION DATA:
25     : APPLICATION NUMBER: US/09/308,207
26     : FILING DATE: 13-May-1999
27     : CLASSIFICATION: <Unknown>
28     :
29     : PRIOR APPLICATION DATA:
30     : APPLICATION NUMBER: 60/030,601
31     : FILING DATE: 13-NOV-1996
32     : ATTORNEY/AGENT INFORMATION:
33     : NAME: Glaister, Debra
34     : REGISTRATION NUMBER: 33,888
35     : REFERENCE/DOCKET NUMBER: GC 369-2
36     : TELECOMMUNICATION INFORMATION:
37     : TELEPHONE: 650-864-7620
38     : TELEFAX: 650-845-6504
39     : TELEX: <Unknown>
40     :
41     : INFORMATION FOR SEQ ID NO: 59:
42     : SEQUENCE CHARACTERISTICS:
43     : LENGTH: 607 amino acids
44     : TYPE: amino acid
45     : STRANDEDNESS: single
46     : TOPOLOGY: linear
47     :
48     : MOLECULE TYPE: NO: US2003002232A1e
49     : SEQUENCE DESCRIPTION: SEQ ID NO: 59:
50     :
51     :
52     :
53     :
54     :
55     :
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :
101    :
102    :
103    :
104    :
105    :
106    :
107    :
108    :
109    :
110    :
111    :
112    :
113    :
114    :
115    :
116    :
117    :
118    :
119    :
120    :
121    :
122    :
123    :
124    :
125    :
126    :
127    :
128    :
129    :
130    :
131    :
132    :
133    :
134    :
135    :
136    :
137    :
138    :
139    :
140    :
141    :
142    :
143    :
144    :
145    :
146    :
147    :
148    :
149    :
150    :
151    :
152    :
153    :
154    :
155    :
156    :
157    :
158    :
159    :
160    :
161    :
162    :
163    :
164    :
165    :
166    :
167    :
168    :
169    :
170    :
171    :
172    :
173    :
174    :
175    :
176    :
177    :
178    :
179    :
180    :
181    :
182    :
183    :
184    :
185    :
186    :
187    :
188    :
189    :
190    :
191    :
192    :
193    :
194    :
195    :
196    :
197    :
198    :
199    :
200    :
201    :
202    :
203    :
204    :
205    :
206    :
207    :
208    :
209    :
210    :
211    :
212    :
213    :
214    :
215    :
216    :
217    :
218    :
219    :
220    :
221    :
222    :
223    :
224    :
225    :
226    :
227    :
228    :
229    :
230    :
231    :
232    :
233    :
234    :
235    :
236    :
237    :
238    :
239    :
240    :
241    :
242    :
243    :
244    :
245    :
246    :
247    :
248    :
249    :
250    :
251    :
252    :
253    :
254    :
255    :
256    :
257    :
258    :
259    :
260    :
261    :
262    :
263    :
264    :
265    :
266    :
267    :
268    :
269    :
270    :
271    :
272    :
273    :
274    :
275    :
276    :
277    :
278    :
279    :
280    :
281    :
282    :
283    :
284    :
285    :
286    :
287    :
288    :
289    :
290    :
291    :
292    :
293    :
294    :
295    :
296    :
297    :
298    :
299    :
300    :
301    :
302    :
303    :
304    :
305    :
306    :
307    :
308    :
309    :
310    :
311    :
312    :
313    :
314    :
315    :
316    :
317    :
318    :
319    :
320    :
321    :
322    :
323    :
324    :
325    :
326    :
327    :
328    :
329    :
330    :
331    :
332    :
333    :
334    :
335    :
336    :
337    :
338    :
339    :
340    :
341    :
342    :
343    :
344    :
345    :
346    :
347    :
348    :
349    :
350    :
351    :
352    :
353    :
354    :
355    :
356    :
357    :
358    :
359    :
360    :
361    :
362    :
363    :
364    :
365    :
366    :
367    :
368    :
369    :
370    :
371    :
372    :
373    :
374    :
375    :
376    :
377    :
378    :
379    :
380    :
381    :
382    :
383    :
384    :
385    :
386    :
387    :
388    :
389    :
390    :
391    :
392    :
393    :
394    :
395    :
396    :
397    :
398    :
399    :
400    :
401    :
402    :
403    :
404    :
405    :
406    :
407    :
408    :
409    :
410    :
411    :
412    :
413    :
414    :
415    :
416    :
417    :
418    :
419    :
420    :
421    :
422    :
423    :
424    :
425    :
426    :
427    :
428    :
429    :
430    :
431    :
432    :
433    :
434    :
435    :
436    :
437    :
438    :
439    :
440    :
441    :
442    :
443    :
444    :
445    :
446    :
447    :
448    :
449    :
450    :
451    :
452    :
453    :
454    :
455    :
456    :
457    :
458    :
459    :
460    :
461    :
462    :
463    :
464    :
465    :
466    :
467    :
468    :
469    :
470    :
471    :
472    :
473    :
474    :
475    :
476    :
477    :
478    :
479    :
480    :
481    :
482    :
483    :
484    :
485    :
486    :
487    :
488    :
489    :
490    :
491    :
492    :
493    :
494    :
495    :
496    :
497    :
498    :
499    :
500    :
501    :
502    :
503    :
504    :
505    :
506    :
507    :
508    :
509    :
510    :
511    :
512    :
513    :
514    :
515    :
516    :
517    :
518    :
519    :
520    :
521    :
522    :
523    :
524    :
525    :
526    :
527    :
528    :
529    :
530    :
531    :
532    :
533    :
534    :
535    :
536    :
537    :
538    :
539    :
540    :
541    :
542    :
543    :
544    :
545    :
546    :
547    :
548    :
549    :
550    :
551    :
552    :
553    :
554    :
555    :
556    :
557    :
558    :
559    :
560    :
561    :
562    :
563    :
564    :
565    :
566    :
567    :
568    :
569    :
570    :
571    :
572    :
573    :
574    :
575    :
576    :
577    :
578    :
579    :
580    :
581    :
582    :
583    :
584    :
585    :
586    :
587    :
588    :
589    :
590    :
591    :
592    :
593    :
594    :
595    :
596    :
597    :
598    :
599    :
600    :
601    :
602    :
603    :
604    :
605    :
606    :
607    :
608    :
609    :
610    :
611    :
612    :
613    :
614    :
615    :
616    :
617    :
618    :
619    :
620    :
621    :
622    :
623    :
624    :
625    :
626    :
627    :
628    :
629    :
630    :
631    :
632    :
633    :
634    :
63
```

	Query Match	5.7%; Score 116; DB 11; Length 607;
	Best Local Similarity	24.1%; Pred. NO.0.083;
	Matches	71; Conservative 42; Mismatches 112; Indels 70; Gaps 15
QY	ATPAQTATKADADLARKYPELVAAEVSROEDAVATVAKRSAGGVKAA--QAAIKSAGIN	177
	: : : : : : : : : : : : : : : : : :	
DG	12 ATTEVALA---SDYQARAFVASGIVA----TTGGKKGRDRLIAGTLLALEGLAKTPWMS	64
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	178 LNRGR-----TPIPGSGTGQSKEVSEGTLLNAGDTVTATTIRQTNEMTVYNVTQSASEVMKL	233
	: : : :	
DG	65 SDYSRIYLNEAPRVPVGDMATEITTE-TIIT-ESTMIGHNPQT-PGSGVGY-GVGSTTIALL	118
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	234 RR-----QIEGKTLAAD-----GVIAVGIKPEPDGVIVYEKKR	266
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
DG	119 GRATLPAAQVYAEIGIVILIDDAVDFLDAVWMLEALDRGINVAAILIKKDQGVIV--NNR	176

QY 267 LIFADVVVNESTGQITTLRAAVND---QNIIMGLYKRVIMDDVAADNAF-----313
 Db 177 LRKTLFVVE-----VTLLEQVEGMAAIEVAAPQVVRILSNPYGIATPFGLSPESTQA 232
 QY 314 VVRQQAIVTGAKDITWIVVAAGSGEEREVTAQAQGGITWITISGLKGGDKVYNG 368
 Db 233 IVDIAALLIGNSAVVLKTPGGDVDSKVIPIAG---NLVYSGEKRGGEADVAG 282

RESULT 24

```

US-10-156-761-13234
Sequence 13234, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13234
LENGTH: 340
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13234

```

```

Query Match      5.7%; Score 115; Length 340;
Best Local Similarity 21.8%; Pred. No. 0.043;
Matches 87; Conservative 56; Mismatches 129; Indels 128; Gaps 19;

QY   MRAAAALAAVALVLSGCGKGDAAGQPAPGREAPAFVGVTVHPQTVALTELPGRL E 68
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
DB   LRAAAYTGAVALTLTLACGGSGSGSGSGSG--SGSVKVGILIT-----KTDTNPFVK 62

QY   SIATADVRAOVGGIIOQRLLFOEGSYTFAGOPLIQIDISTTEANLESAPAQLATQATLAK 128
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
DB   MKEGAETKAKEKV-----OLSTA---AGK 84

QY   ADDDLARRYRKLVAEAENVSRQDYDAAYTA-KRSAEAGYKAQAQAIKSAGINMRITAPI 187
DB   FDDDNA-----GOYVAIEIMWYAAGVK-----GLITTPDSKAIIV 118

QY   SGFTIGOSKVESEGLTNAGDTTVALATTIRQTNPMVNVNTQSASEVKLRROIAE----- 239
DB   PAL-OKARKKGYLVIALDT-----PTPEESAVNDLPFIDNLKAGOLLGEYKAWMG 169

QY   -GKLLAADGV--IANGIK---FDDGTYYPEKGRILLFADPVNNESTGQITLRAAVENDQ 291
DB   KKAKIALDLAPGVSVGVRHNGFLKGFATDRDVCVCAQDTGGDAQGTAM-----E 222

QY   NIIIM-PGLYRVRLMDQVAVDNFAFYVPOQAVTGADTWNIVNAOG-----MEPREVT 343
DB   NCICQAKPGINVVYTIINEPALGAYTALK--AGRGREDVLIVSDGCGTQAVAKDKIA 279

QY   VAAQQ-----GTNWIVTSGLKDGDKKVNVVEGISLAGIT 375
DB   ATSQOYPLKMAAGCVKAAVITYA-KDGKK--ASGYTTTGVT 316

RESULT 25
US-10-156-761-13251
; Sequence 13251, Application US/10156761
Publication No. US20030119018A1

```

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13251
LENGTH: 1208
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13251

Query Match 5.7%; Score 114.5; DB 15; Length 1208;
Best Local Similarity 21.7%; Pred. No. 0.31;
Matches 95; Conservative 53; Mismatches 170; Indels 119; Gaps 21;

OY 11 AALAAVALVSSGKGDAAGQPRAPAVGVVYHPTVALTVLELPGRLSL 70
DB 10 AALISAGIVFALLPVGRAA-ARDAGTPSPPOASTARTVTLVTDPTVT- DLPGKKTV 67
OY 71 -----RTADVRAOV-----GGIIOKRLFQSGVYRAGOPTLQIDSS 106
DB 68 TVERPRGATGAVRTQVGVGASVVDLEALPYLRAGTLRRLFDVSAALRQGIS- DRK 123
OY 107 TYEANL-----ESARAQI-ATAQATL-----AKADDLAR--YKPLVAEAVSRQED 151
DB 124 TGEFPLIYGVGKAAVTPSGATRRSLTSGAAVADADKSTFWRSWTRQGIHWMID 183
OY 152 AAV---TKRBAEAGVRA-----QAIRKAGINLRKRTAIS---GRIGSK 195
DB 184 ARVEDAMESNAQIGTRAMADAGLTGDEGVAVVLDVTGDTTHPDLAGRVSRSKSTIDEE 243
OY 196 VSEGLTNACDPTVLATI-----ROTNPMYVAVTQASAEVVK-----LRQIAE 239
DB 244 VAD---RNGHGHVTSYVGGSGAASDGTGRGAP---GATLAVGVILSDQAGSESQIIA 297
OY 240 GKLLAADGVIAVGIKFDDGTVPKGRLLFADPV--VNESTGQITLRAAVPNDQNIIMP 297
DB 298 GMEWAAARVRAIRIVMSLSTGTEASDGTDPMAEAVDTLSEETG-----A 340
OY 298 LTVRLVLMQVAVDNAFVVPQAVTRGADDTWIVVNAOGGMEPREVTVAOQOGTMMIVTSG 357
DB 341 LRV-----VVAAGNT-GAPSSIGSPGADSAITVGA-----VDSRAAAYFTSG 383
OY 358 LKDGKVVVEGISINGI 374
DB 384 PRHGDNALKPDIAAPGV 400

RESULT 26
US-09-815-242-5883
Sequence 5883, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5883
LENGTH: 837
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5883

Query Match 5.6%; Score 114; DB 9; Length 837;
Best Local Similarity 23.3%; Pred. No. 0.2;
Matches 82; Conservative 43; Mismatches 139; Indels 88; Gaps 19;

OY 112 LESARAOATATQATLAKADADLARKP---LVAAEAVSRQEDYAAVTAKSAAEAG----- 163
DB 18 VEHATLELNTMTLTKALADKATKASGVVNAADAKRQAYDEKVTAAENIVGTPTPT 77
OY 164 -----VKAQAKISAGINLN-----RSRTAPISGFIG-----OSKXSEGLLNAG 205
DB 78 LTPADVNTAAQOVNTAKTOLNGLNLEAVAKONANTALDGLTSLNGPOKAKKE-----QVG 133
OY 206 DTVLATITROTNPMYVAVTQASAEVVK-LRQIAEGLKLLADGVIAVGIKFDDGTVPPE- 263
DB 134 CATLPNVQYTR---DNAQTLNTAMKGLRDSIAN-----EATIKAGQNTDASQNKQT 183
OY 264 --KGRLLFADPVNESTG-----QITLRAAVPNDQNIIMPGLVYRVLMQVA 308
DB 184 DYNSAVTAARAKIIIGQTPSPNNAQEIINQAKQVTAQOALNGQENL-----RTA 232
OY 309 VDNA-FVVPQAVTRGAKDTWIVVNAOGGMEPREVTVAOQ--OGTMMIVTSGKDG--- 361
DB 233 QTNAKQHLNGSLDITLDAQDAVK-RQIEGATHVNEVTOACNNADALNTAMTN-LKNGIQD 290
OY 362 DKVVEGISINGITGAKK-----VTPKEMASSENOAAAP---OSGVOTASE 404
DB 291 QNTIKQGVNFTDADEAKRNAYTNAVTOAE-QIINKAQGPNTSKQGVETALE 340

RESULT 27
US-09-815-242-13080
Sequence 13080, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242

Query Match	5.6%;	Score 114;	DB 9;	Length 875;
Best Local Similarity	23.3%;	Pred. No. 0.22;		
Matches	82;	Conservative	43;	Mismatches 139;
			Indels	88;
			Gaps	19;

RESULT 28
 US-09-815-242-5835
 : Sequence 5835 Application US/09815242
 : Patent No. US20020061569x1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Karl I.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : TITLE OF INVENTION: Prokaryotes
 : FILE REFERENCE: ELITRA.011a
 : CURRENT APPLICATION NUMBER: US/09/815,242
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR APPLICATION NUMBER: 60/191,078
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR APPLICATION NUMBER: 60/206,848

Query Match	5.6%;	Score 114;	DB 9;	Length 2434;
Best Local Similarity	23.3%;	Pred. No. 0.99;		
Matches	82;	Conservative	43;	Mismatches 139;
			Indels	88;
			Gaps	19;

RESULT 29
 US-09-815-242-12996
 ; Sequence 12996, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 12996
LENGTH: 6281
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match 5.6%; Score 114; DB 9; Length 6281;
Best Local Similarity 23.3%; Pred. No. 4;
Matches 82; Conservative 43; Mismatches 139; Indels 88; Gaps 19;

112 LESARAQATATQAATLAKADADLARYK---LVAAEVSRQEYDAVATKRSNAAG----- 163
1399 VEHTATELNTMTLAKRAIADAEFKASGVNADANKRQAVDEKVTAAENIVSGTPTPT 1458
164 -----VKAAQAAIKSAGINLN-----RSRTIAPISGFTG-----QSKVSEGLLNAG 205
1459 LTPADVNTAAQVNTAKTQNLGNHNI, EVAKQANATAIDGLTSLNGPOKAKLKE-----QVG 1514
206 DTVYATIRQTNPMVNTQSGSEVWK-LRROIAEGKLLADGVAVGKFPDGTVYE- 263
1515 QATLTPNVQTVR---DNAQTLNTAMKGLRDSIAN-----EATLKAQONTVDASONKOT 1564
264 --KGLLPADPVNNESTG-----QITLRAAVPNDQNLMPGLYVRVLMDOVA 308
1565 DYNNAVTAAKAIIGTGTSPSMAQOINQAKQVTKQAALNGEHL-----RTA 1613
309 VDNA--FVVPQQAIVTRGAKDVTMIVNAQSGMEPREVTVAAQ--QGTNMTVTSGLKDG--- 361
1614 QTNNAQHNLGSLDLTDAQDAVK-RQIBGATHVNEVTOQNNAADALNTAMTN-LRNGIOD 1671
362 DKVAVGEISAGITGAKK-----VTPKEMASSENOAAP---QSGVQGTASE 404
1672 ONTKQGVNFDTDADEAKRNVATNAVTOAE-QILNKAQGPNTSKQGVETALE 1721

RESULT 30

US-10-156-761-12370
Sequence 12370, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO: 12370
LENGTH: 272
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12370

Query Match 5.6%; Score 112.5; DB 15; Length 272;
Best Local Similarity 24.8%; Pred. No. 0.052;

Matches 72; Conservative 29; Mismatches 112; Indels 77; Gaps 12;
QY 12 AALAAVALVSSCGKGDAAGGQAPAGREAPVGVVTHPQTVALTYELPRLSLR 71
DB 24 AGVAGGVSTLVAAGAAG-AANAAP-----VQTYELP-----TL 58
QY 72 TADVAQVGGIIOKRLFOEGSYVRAGPLVQIDSTVEA-----NLESARAQATATQA 127
DB 59 TADLATQVA-----QSADATQAAANYELQERDAAAQAKQA 97
QY 128 KADADLARYKPLVAEVSROEYDAVATKRSNAAGKAAQAAIKSAGININRRTAP- 186
DB 98 KSDIADAKKKAATKKAAPAAARRAAERASASERATLSAS-ASTGTSTSTSTATG 156
QY 187 -----ISGFQSGKVSSEGLTNAGDTT--VLATIRQTNPMVNT--QSGSEVWKLR 234
DB 157 SAAVIAFVRKQIDAVYSGGTGPNMDSGLVQAARKQVSVSLPRSSQDSTAGTVSL 216
QY 235 QIAEGKLL--AADGVIAVGIRKFDGTVPKGRLLPADPVNESTGQI 281
DB 217 SNLQPGDILVYGAGASAVHGVVYGDG-----MFGV-AQNPSTGVV 256

RESULT 31

US-09-815-242-10096
Sequence 10096, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 10096
LENGTH: 405
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10096

Query Match 5.6%; Score 112.5; DB 9; Length 405;
Best Local Similarity 22.2%; Pred. No. 0.094;
Matches 90; Conservative 54; Mismatches 142; Indels 119; Gaps 18;

QY 15 AAVALVSSCGKGDAAGGQAPAGREAPVGVVTHPQTVALTYELPRLSLRAD 74
DB 14 SVAQVATVTHKKKGDAV-----VRDEVLVEITDKVLEVP----- 50
QY 75 VRAQVGGIIOKRLFOEGSYVRAGPLVQI--DSSTVEANLES-----ARAQATATQA- 123

Db 51 --ASADGILDVLEDEBGTVTTSRQILGRLREGNSAGKETSASKEKASTPAQOASIEE 108
Qy 124 -----ATLAKADADLARKYPLVAEAHSVQEYD---AAVTAKRSA-EAGVKAQ 168
Db 109 QNNDALSPAIRRLAEHNLDAISAKGTGVRGLTREDVEKHLAKAPAKESAPAAAAPAAQ 168
Qy 169 AAIKAGINLRSRRTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMVAVNTQSAS 228
Db 169 PALAA-----RSEKRVPMTRL--RKRVAERLLEAKNSTAMLTTFNEVN-----MK 211
Qy 229 EVMKLRQIAEGKLLAADGVIAVGIKPDDGTV-----YPEKGRLLFPDPVNVNЕСТQOIT 282
Db 212 PIMDLRKQYGEA-FEKRHGIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYHNYFDVS 270
Qy 283 LRAAVPNDQNLIMPGLYVRVLMQVAVDNAFVVPQOAVTRGAKD-----TVMIVNAQCGM 337
Db 271 MAVSTPR-----GLVTPVLRD---VD-----TLGMADIEKKIKELAVAGRGD- 309
Qy 338 EPREVTVAQOQGTWMTVTSGLKDGDKVVEGISIAGITGAKKVP 382
Db 310 ---KLTVEDLTGNGFTITNG-----GVFGSLMSTP 336

RESULT 32

US-10-078-107-6
; Sequence 6, Application US/10078107
; Publication No. US20020182688A1
; GENERAL INFORMATION:
; APPLICANT: IZUI, HIROSHI
; APPLICANT: HARA, YOSHITAKO
; APPLICANT: SATO, MASAAZU
; APPLICANT: AKIYOSHI, NAOKI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID
; FILE REFERENCE: 219846USO
; CURRENT APPLICATION NUMBER: US/10/078, 107
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: JP 2001-044134
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 405
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-078-107-6

Query Match 5.6%; Score 112.5; DB 14; Length 405;

Best Local Similarity 22.2%; Pred. No. 0.094;

Matches 90; Conservative 54; Mismatches 142; Indels 119; Gaps 18;

Qy 15 AAAVALVLSGCGGDAAGGQAPAGREAPVVGVTVAHPQTVALTVELPGRLESIRTD 74
Db 14 SVAATATVATMHKKRGDAV-----VRDEVLVEITDKVLEVP-----50
Qy 75 VRAOVGGIIQKRLFOEGSYRAGOPLYOI---DSSTYEANLES-----ARAQIATAQ- 123
Db 51 --ASADGILDVLEDEBGTVTTSRQILGRLREGNSAGKETSASKEKASTPAQOASIEE 108
Qy 124 -----ATLAKADADLARKYPLVAEAHSVQEYD---AAVTAKRSA-EAGVKAQ 168
Db 109 QNNDALSPAIRRLAEHNLDAISAKGTGVRGLTREDVEKHLAKAPAKESAPAAAAPAAQ 168
Qy 169 AAIKAGINLRSRRTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMVAVNTQSAS 228
Db 169 PALAA-----RSEKRVPMTRL--RKRVAERLLEAKNSTAMLTTFNEVN-----MK 211
Qy 229 EVMKLRQIAEGKLLAADGVIAVGIKPDDGTV-----YPEKGRLLFPDPVNVNЕСТQOIT 282
Db 212 PIMDLRKQYGEA-FEKRHGIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYHNYFDVS 270
Qy 283 LRAAVPNDQNLIMPGLYVRVLMQVAVDNAFVVPQOAVTRGAKD-----TVMIVNAQCGM 337
Db 271 MAVSTPR-----GLVTPVLRD---VD-----TLGMADIEKKIKELAVAGRGD- 309

Qy 338 EPREVTVAQOQGTWMTVTSGLKDGDKVVEGISIAGITGAKKVP 382
Db 310 ---KLTVEDLTGNGFTITNG-----GVFGSLMSTP 336

RESULT 33

US-10-077-751-6
; Sequence 6, Application US/10077751
; Publication No. US20020192772A1
; GENERAL INFORMATION:
; APPLICANT: SATO, MASAAZU
; APPLICANT: AKIYOSHI, NAOKI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-GLUTAMIC ACID
; FILE REFERENCE: 219849USO
; CURRENT APPLICATION NUMBER: US/10/077, 751
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: JP 2001-044135
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 405
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-077-751-6

Query Match 5.6%; Score 112.5; DB 14; Length 405;

Best Local Similarity 22.2%; Pred. No. 0.094;

Matches 90; Conservative 54; Mismatches 142; Indels 119; Gaps 18;

Qy 15 AAAVALVLSGCGGDAAGGQAPAGREAPVVGVTVAHPQTVALTVELPGRLESIRTD 74
Db 14 SVAATATVATMHKKRGDAV-----VRDEVLVEITDKVLEVP-----50
Qy 75 VRAOVGGIIQKRLFOEGSYRAGOPLYOI---DSSTYEANLES-----ARAQIATAQ- 123
Db 51 --ASADGILDVLEDEBGTVTTSRQILGRLREGNSAGKETSASKEKASTPAQOASIEE 108
Qy 124 -----ATLAKADADLARKYPLVAEAHSVQEYD---AAVTAKRSA-EAGVKAQ 168
Db 109 QNNDALSPAIRRLAEHNLDAISAKGTGVRGLTREDVEKHLAKAPAKESAPAAAAPAAQ 168
Qy 169 AAIKAGINLRSRRTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMVAVNTQSAS 228
Db 169 PALAA-----RSEKRVPMTRL--RKRVAERLLEAKNSTAMLTTFNEVN-----MK 211
Qy 229 EVMKLRQIAEGKLLAADGVIAVGIKPDDGTV-----YPEKGRLLFPDPVNVNЕСТQOIT 282
Db 212 PIMDLRKQYGEA-FEKRHGIRLGFMSFYKAVVEALKRYPEVNASIDGDDVYHNYFDVS 270
Qy 283 LRAAVPNDQNLIMPGLYVRVLMQVAVDNAFVVPQOAVTRGAKD-----TVMIVNAQCGM 337
Db 271 MAVSTPR-----GLVTPVLRD---VD-----TLGMADIEKKIKELAVAGRGD- 309
Qy 338 EPREVTVAQOQGTWMTVTSGLKDGDKVVEGISIAGITGAKKVP 382
Db 310 ---KLTVEDLTGNGFTITNG-----GVFGSLMSTP 336

RESULT 34

US-09-741-669-414
; Sequence 414, Application US/09741669
; Patent No. US2002022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741, 669
; CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: US 60/173005
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 481
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 414
LENGTH: 542
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-669-414

Query Match 5.5%; Score 112; DB 9; Length 542;
Best Local Similarity 24.2%; Pred. No. 0.16; Indels 64; Gaps 16;
Matches 65; Conservative 47; Mismatches 93;

QY 168 QAKISAGINLRSHITAPI--SGFIGOSKVSSEGT-----LLNAGD--TVLATIR 214
DB 113 ENGISAEMIDPOCAHIEBNAVPALIGAVKVPDGVDPRLTAAMMDAKHGAIVLTAH 172
QY 215 QTNPMYVNTGASAEV-MKLRRQI-AEGKLLADGVI-AVGI-----KEDDGTV--YPE 263
DB 173 EVTGL---IREGATVGVVRNHLTGCTQALHAPVYVNAAGIWGHIAEVADLRIMFPA 229
QY 264 KGRLLFADPVNYESGQITLRAAVNDONILMPGLV-----RYLMD 305
DB 230 KSLILIMHRIHQ---HYINRCRKSDDAIVLPGDTISLIGTSLRIDYNEIDNRVTAE 286
QY 306 QYAV---DNAFVPPQAAVTRGAKDTMTVNAAGMEPREVTVAQQGNTWITVSGLKDG 362
DB 287 EVDILRREGEKAPVMATR-----ILKAYSGRPLVASDDDSGN--VSRGIVLDD 337
QY 363 KVVVEGI-SIAGITAKKVTPK--EWAS 387
DB 338 HAERDGLDGFITITGGKLTMTYRLAMEWAT 366

RESULT 35

US-09-159-469-50
Sequence 50, Application US/09159469
Patent No. US20020064535A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: THERAPY OF EHRLICHIA INFECTION
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/159,469
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/106,582
FILING DATE: 29-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:
LENGTH: 578 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-159-469-50

Query Match 5.5%; Score 112; DB 9; Length 578;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 98; Conservative 64; Mismatches 151; Indels 164; Gaps 18;

QY 6 FKAMRAAA---LAAVALVSSCGKGDAAQGGPAGREAPAPVGVVTHPQTALVTE 62
DB 134 FKIKSAAEHPQATDPVSVEASATSGVD--TCEBOEIOEAPV----- 174
QY 63 LRGLESRTADVAAQVGGIQRKLPQGSVVRAGOPLOYI----- 103
DB 175 ---IEVEYTE-----QEVILEBGTILIDEPVAAQVPVVAEALPGVEAAEAI 221
QY 104 ---DSSTVEA-----NLESARQLATAQ---ATLAKADADLARVPLVAAEA---VSR 147
DB 222 SLEENKQOEVVVAPEAQLESAPESVAPQPESTVLGVAEGD---KSEVSVEANADVAQ 278
QY 148 QEV-----DAVTAKRSAEA-----GVK 165
DB 279 KEVISGOOEQIEAIEALEGTEAPVEKEETEVLKEDTLIDLEQVAAQVPVVAEALPGVE 338
QY 166 AQAIAISAGINLRSHITAPISGFIGOSK-----VSRGTLINADDTVLA 211
DB 339 AAEAIIVPLEENKQOEVVVAPEAQLESAPESVAPQPESTVLGVAEGD---KSEVSVEANADVAQ 398
QY 212 TIROTNPYVNTGASAEV-MKLRRQI-AEGKLLADGVI-AVGI-----KEDDGTV--YPE 263
DB 399 GMQOEAGISDQETATEEVEVE-----VSVETKEBPEVILEBGTILIDLE 444
QY 272 -----PVNES--TGQITLRAAVNDONILMPGLYRVLMDOVAVDNAFVPPQO----- 318
DB 445 QPVAQVPEVAEALPGVEAAEAIIVPLEENKQOEVVVAPEAQLESAPESVAPQPESTV 504
QY 319 -AVTRGAKDTMTVNAAGMEPREVTVAQQGNTWITVSGLKDDKVVVEGISTAGI 374
DB 505 LGVTGDLKSEVSVEADAGMQ-QEAGISDDE-----TQATEEVEKVSVEADAGM 554

RESULT 36

US-09-798-042-50
Sequence 50, Application US/09798042
Patent No. US20020068343A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
FILE REFERENCE: 210121.439C7
CURRENT APPLICATION NUMBER: US/09/798,042
CURRENT FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 108
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 50
LENGTH: 578
TYPE: PRT
ORGANISM: Ehrlichia sp.
US-09-798-042-50

Query Match 5.5%; Score 112; DB 9; Length 578;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 98; Conservative 64; Mismatches 151; Indels 164; Gaps 18;

QY 6 FKAMRAAA---LAAVALVSSCGKGDAAQGGPAGREAPAPVGVVTHPQTALVTE 62
DB 134 FKIKSAAEHPQATDPVSVEASATSGVD--TCEBOEIOEAPV----- 174
QY 63 LRGLESRTADVAAQVGGIQRKLPQGSVVRAGOPLOYI----- 103
DB 175 ---IEVEYTE-----QEVILEBGTILIDEPVAAQVPVVAEALPGVEAAEAI 221
QY 104 ---DSSTVEA-----NLESARQLATAQ---ATLAKADADLARVPLVAAEA---VSR 147
DB 222 SLEENKQOEVVVAPEAQLESAPESVAPQPESTVLGVAEGD---KSEVSVEANADVAQ 278
QY 148 QEV-----DAVTAKRSAEA-----GVK 165
DB 279 KEVISGOOEQIEAIEALEGTEAPVEKEETEVLKEDTLIDLEQVAAQVPVVAEALPGVE 338
QY 166 AQAIAISAGINLRSHITAPISGFIGOSK-----VSRGTLINADDTVLA 211
DB 339 AAEAIIVPLEENKQOEVVVAPEAQLESAPESVAPQPESTVLGVAEGD---KSEVSVEANADVAQ 398
QY 212 TIROTNPYVNTGASAEV-MKLRRQI-AEGKLLADGVI-AVGI-----KEDDGTV--YPE 263
DB 399 GMQOEAGISDQETATEEVEVE-----VSVETKEBPEVILEBGTILIDLE 444
QY 272 -----PVNES--TGQITLRAAVNDONILMPGLYRVLMDOVAVDNAFVPPQO----- 318
DB 445 QPVAQVPEVAEALPGVEAAEAIIVPLEENKQOEVVVAPEAQLESAPESVAPQPESTV 504
QY 319 -AVTRGAKDTMTVNAAGMEPREVTVAQQGNTWITVSGLKDDKVVVEGISTAGI 374
DB 505 LGVTGDLKSEVSVEADAGMQ-QEAGISDDE-----TQATEEVEKVSVEADAGM 554

Db 134 FKIEKSAEFPOTVDSVSESATGSGVD--TQEEQETIDQEPAP----- 174
Qy 63 LFGRLSELTADVRAQVGIIQKRLFOEGSYVRAQGPLOYI----- 103
Db 175 ---LEEVETEE-----QEVILEBETLIDLPQVAVVVAEALPGVEAAEIVP 221
Qy 104 ---DSSTTEA-----NIESARAQALATQ---ATLAKADADLARYKPLVAAEA---VSR 147
Db 222 SLEENKQEVVVAPEAQOLESAPESAPAPQPESTVLGVAEGLD---KSEVSVEANADVAQ 278
Qy 148 QEY-----DAAVTAKRSAA-----GVK 165
Db 279 KEVISGQOEQETIAELTEGTAPEVEKETETVILKEDTLIDLEQPAQVPAVVAEALPGVE 338
Qy 166 AAQAAIKSAGININRSRITAPISGFIGOSK-----VSEGLTNAGDTYLA 211
Db 339 AAEEAVPLEENKQOEVVVAPEAQOLESAPESAPAPQPESTVLGVTEGDLKSEVSVEADA 398
Qy 212 TIRQTNPMYVAVTOSASEVMKLRQIAGKLLAADGVIAVGIKFDDGTVPYPERKRLFPAD 271
Db 399 GMOQFAGISDQETQATEEVEKE-----VSVEKTEPEVILIEGTLIDLE 444
Qy 272 ---PVNSES---TGQTLTAAVNDQNIIMPLGLYVAVLMDQVAVNDNAFVVPQ----- 318
Db 445 QPVAQVPAVAAEELPGVEAAEIVPSLEENKQOEVVVAPEAQOLESAPESAPAPQPESTV 504
Qy 319 -AVTRGADTVMIVAAQGMPEREVTVAAQOQGTNMTVSGLKDGDKVVEGISIAGI 374
Db 505 LGVTEGDLKSEVSVEADAGM-QEAGISDQE-----TQATEEVEKVEVSVEADAGM 554

RESULT 37

US-09-815-242-5885
; Sequence 5885, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykend, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5885
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5885

Query Match 5.5%; Score 111; DB 9; Length 1029;
Best Local Similarity 21.2%; Pred. No. 0.51;

Matches 84; Conservative 58; Mismatches 130; Indels 124; Gaps 19;
Qy 58 ALTVELPERLESLRTADVRAQVGIIQ---KRLFOEGSYVRAQGPLOYIDSTYEANE 113
Db 35 ARAIDALPHLNNAQADVAKSKINAAASINAGVNTVQOQT-----DLNTMGNLQ 83
Qy 114 SARAQALATQATLAKADADLARYKPLVAAEAVSROEYDAATAKRSAAE-GVKAQAAI- 171
Db 84 GA---INDEQTLT-----NSQNVQDATPSKKTATVNAVQAADILN 121
Qy 172 KSAGININRSRITAPISGFIGOSKVSSEGLT-----NAGDTYLATIRQTNPMYV 221
Db 122 KSNQNKTKQDVTEAMQVNSAKNNLDQTRLLDQAKQAKOQLNMNMTLTTAQKTN----- 177
Qy 222 NVTOSASVMTLRQIAGKLLAADGVIAVGIKFDDGTVPYPERKRLFPADVNESTQI 281
Db 178 ---LTVQINSQTTVA-----GVQ---TVQSNANTL---DQAMN----- 206
Qy 282 TLRAAVNDQNIIMPLGLYVAVLMD-QVAVNDNAFVVPQAVTRGADTVMIVAAQGMPE- 339
Db 207 TLRSINKKATKASQEDVYDANDKQATYNNAAV-----AAETIINANSNPEMPS 257
Qy 340 -----REVTVAAQOQGTNMTVTVSGLKDGDKVVEGISIAGITGAKK 379
Db 258 TITQKAEQVNSKTKALNGDENILTAQAKAKTYLNTLITSDAQK---NNLISQITTSATR 313
Qy 380 V---TPKEWASSENQA-AAPOSGVQTAQSEAKTASE 410
Db 314 VSGVDTVKQNAQHLQDQAMASLQNGINNESQVKSSEK 349

RESULT 38

US-09-815-242-13083
; Sequence 13083, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykend, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13083
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-13083

Query Match 5.5%; Score 111; DB 9; Length 1048;
Best Local Similarity 21.2%; Pred. No. 0.53;

Matches 84; Conservative 58; Mismatches 130; Indels 124; Gaps 19;

QY 58 ALTYELPERLSIRADYRAOVGIIQ-----KRLFOEGSYRAGOLYQIDISTYEANLE 113
 DB 47 ARADIDALPHLNNAQADYKSKINAASNTAGVNTVQOGCT-----DLNTMGMLQ 95
 QY 114 SARAQLATAQATLAKADADLARYPELVAAEAVSROEYDAAVTAKSAAE-GVKAQAQAI- 171
 DB 96 GA-----INEQOTTL-----NSQNYQDAIPSKKTATATNVAQAADILN 133
 QY 172 KSAGINLNRSRITAPISGFIGOSKVSSEGLT-----NAGCTVTLATITROTNPVY 221
 DB 134 KSMGNKTKDQDYTEMENQVNSAKNNLDGIRLIDQAKQOTAKOQUNMTHLTAAOKIN----- 189
 QY 222 NVTQASAVMKLRQIAGKGLLAADGVIAVGKFPDGTVYPEKRLFPADPVNESTQOI 281
 DB 190 -----LITQINSGITVA-----GVQ-----TVQSNANTL---DQAMN----- 218
 QY 282 TLRAAVPNDONILMGLYVRVLM-DQAVDNAFVVPQOAVTRGAKDTWIVNAQGMGP- 339
 DB 219 TLROSIANKDATKASSEDYVDANDKQTAVNNNAV-----AAEETIINNNSNPNMPS 269
 QY 340 -----REYTVAAQOGCTMWIT-SEGLKDGKVVVEGISTAGITGAKK 379
 DB 270 TITQKAEQVNSKTLANGDENILTAQONAKTYLANTLITSIDAQ-----NNLISQITSATR 325
 QY 380 V-----TPKEMASSENQ-AAPQSGVQTASEAKTASE 410
 DB 326 VSGVDTVRKQNAQHLQDQAMASLQNGINNESQVSKSEK 361

RESULT 39

US-09-782-714-1
 ; Sequence 1, Application US/09782714
 ; Patent No. US20020034808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dazet, Seth A
 ; APPLICANT: Zhang, Gongyi
 ; APPLICANT: Campbell, Elizabeth
 ; APPLICANT: Minkin, Leonid
 ; APPLICANT: Severinov, Konstantin
 ; TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS
 ; FILE REFERENCE: 600-1-258
 ; CURRENT FILING DATE: 2001-02-13
 ; PRIOR APPLICATION NUMBER: US/09/782,714
 ; PRIOR FILING DATE: 1999-09-15
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1525
 ; TYPE: PRT
 ; ORGANISM: Thermus aquaticus
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1247)
 ; OTHER INFORMATION: Any amino acid can be at this position
 US-09-782-714-1

Query Match 5.5%; Score 111; DB 9; Length 1525;
 Best Local Similarity 19.8%; Pred. No. 0.92;
 Matches 98; Conservative 73; Mismatches 161; Indels 164; Gaps 25;

QY 21 VLSGCGKGDAAQGGQ-----PAGREAPAV-----GVVTVHPQTVALTVELPG 65
 DB 1023 VMAQSGARGNPQIRQLCGMRGLMQPSEGETFEVVRSSFRGLTVLEVFISSHGARKG 1082
 QY 66 RLESRTADVRAQVGGIIQKRLFOE-----GSYVRAQGPLYQIDISTYANLES 114
 DB 1083 ADTALRTID-----SGYLTRKLVDAVHEIVREADCGTTKYISVPLFQWDEVTRTLRLRK 1137
 QY 115 ARAQLATAQATLAKADADLARYPELVAAEAVSROEYDAAVTAKSAAEAGVKAQA--- 169

DB 1138 -----RSDIESGLYGRVLAREVEALGRLEEGRYLSELDVHFLIKAAEAGEV 1184
 QY 170 -----AISKAGINLNRSRITAPIS-----GFIGOSKVSF-GTLL----- 202
 DB 1185 REVERSPPLCTQTRYGVCCQKYGYDLSMAR---PVSIGEAAGVVAASIGEPGTQTLWRT 1241
 QY 203 -NAGDTVTLATITROTNPVYVNTQASAEVVKLRQIAGKGLLAADGVIAVGKFPDGTIVY 261
 DB 1242 FHTGXAAGVDITDGLRVI-----ELFEARRPKKAVISEIDGV---VRIEFG--- 1287
 QY 262 PEKRLFPADPVNESTQOITLRAAVPNDONILMGLYVRVLMQAVDNAFVVPQOAVT 321
 DB 1288 --EDRL-----SVFESEEG-ESKEYKLPRDARLVK-----DGDYVEAGQPLT 1327
 QY 322 RGAKDTWIVNAQO-----GMEPREVTVAAQOGCTMWI--VT----- 355
 DB 1328 RGAIDPHQLLEAKKPEAVERVYLDIETQKVRAGGVKLDKHIETIVRQMLKTYEVTDPGD 1387
 QY 356 SGLKDG-----DKVVVEG-ISA-----GITGAKKVPKEMASSENQAAAPQ 396
 DB 1388 SPLLEGQVLEKMDVEALNERLIAEGKVPVAMKPLLMGVT-KSALSTSKWSL---AASFQ 1442
 QY 397 SGVQTASEAKTASEAE 412
 DB 1443 NTHVLTETEAALAGKRD 1458

RESULT 40

US-10-156-761-7965
 ; Sequence 7965, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMIURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 7965
 ; LENGTH: 7746
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-7965

Query Match 5.5%; Score 111; DB 15; Length 7746;
 Best Local Similarity 20.9%; Pred. No. 10;
 Matches 97; Conservative 65; Mismatches 162; Indels 140; Gaps 20;

QY 41 EAPAPVVGVT-----YHPQTVAL-----TVELPGRLESIRTA 73
 DB 2901 QAAAGVAGVIGIWMAMRHGVLPQTLIHDEPTAHVMSGAVRLLTEPVQMGTTGPRRA 2960
 QY 74 DVRAQVGGIIQKRLFOESYRAGQPLYO-----IDSTYANLES-ARAQLATQA 124
 DB 2961 VSSFQIGGTNAHTIIEEPAAAAEPAAREHNPVVPWVLSKSEVALRAQERLLAFATD 3020
 QY 125 TLKADADLARYPELVAAEAVSROEYDAAVTAKSAAE-AGVKAQAQAIKSGINLNRSR 182
 DB 3021 DVSEVDAGFS-----SATTSALFHRAAVIGSPAEIRAGLELMAAGEPAANVYAGRAH 3074
 QY 183 ITAPISGFI-----GQSKVSEGLTLNAG-----DTTV---LATITROTNPVY 220


```

Db      3075 -SADKVGFLFSGQGSQRLGMGBELYAAYPFPAAYDEVCALLDATVDVAETLHQTS-- 3131
QY      221 VNVTSASEVMKLRROIAGKLLADGV-----IAGIKFDGTYPE 263
Db      3132 ---TOPALFAV-----EVALFRLLSEWGIKRPDIYVAGHSVGEIAAAHVAGVLSLEDA----- 3179
QY      264 KGRLLFADPVNESTGQITLRAAVPNDONILMPGIYVRVMDQYAVDNAFVV---PQQA 319
Db      3180 -AKLVSAFAALMKALPAGGVVAVQAEDDEMLPHL-----IDRAGIAIANGPRSV 3228
QY      320 VTRGAKDTVMIVNAGGMEPREVVAAQOQGTNMIYTS-----GLKDGKYYVEGIS 370
Db      3229 VISGAEDAVTAI-----AEVFAAGKRTSRLKVSFAHSPLMDEPMLLEEFVAVRGLT 3280
QY      371 -----IAGITG--AKVTPKEMASSENOAAAPQSGVOTASE 404
Db      3281 FSEPMIPVSNLTGRLTESYTPPEYVVRHVREAVRFADGVLTLE 3324

```

Search completed: September 8, 2003, 14:07:25
 Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:59:51; Search time 21 Seconds
(without alignments)
830.099 Million cell updates/sec

Title: US-09-889-756a-2
Perfect score: 2019
Sequence: 1 MAFRAFKRMAAALAAVAAL.....AAPSGVQTASEAKTASEAE 412

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfltest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	802.5	39.7	424	US-09-328-352-4199	Sequence 4199, Ap
2	753	37.3	453	US-09-252-991A-16988	Sequence 16988, A
3	698.5	34.6	410	US-09-252-991A-30590	Sequence 30590, A
4	647	32.0	447	US-09-328-352-7383	Sequence 7383, Ap
5	642	31.8	414	US-09-252-991A-27975	Sequence 27975, A
6	451.5	22.4	409	US-09-328-352-4570	Sequence 4570, Ap
7	448	22.2	455	US-09-252-991A-31636	Sequence 31636, A
8	429	21.2	429	US-09-252-991A-31241	Sequence 31241, A
9	393	19.5	502	US-09-252-991A-24388	Sequence 24388, A
10	370.5	18.4	458	US-09-199-637A-149	Sequence 149, App
11	355	17.6	454	US-09-252-991A-17691	Sequence 17691, A
12	277	13.7	394	US-09-252-991A-24485	Sequence 24485, A
13	272.5	13.5	454	US-09-328-352-7897	Sequence 7897, Ap
14	267	13.2	399	US-09-252-991A-20610	Sequence 20610, A
15	256	12.7	406	US-09-252-991A-28543	Sequence 28543, A
16	245.5	12.2	389	US-09-252-991A-21202	Sequence 21202, A
17	237.5	11.8	770	US-09-252-991A-28510	Sequence 28510, A
18	220	10.9	500	US-09-328-352-8183	Sequence 8183, Ap
19	216.5	10.7	367	US-09-328-352-5003	Sequence 5003, Ap
20	215	10.6	487	US-09-252-991A-25447	Sequence 25447, A
21	197.5	9.8	306	US-09-252-991A-19591	Sequence 19591, A
22	184.5	9.1	367	US-09-328-352-7523	Sequence 7523, Ap
23	184	9.1	423	US-09-328-352-6273	Sequence 6273, Ap
24	184	9.1	517	US-09-252-991A-31115	Sequence 31115, A
25	183.5	9.1	354	US-09-252-991A-18501	Sequence 18501, A
26	179	8.9	356	US-09-328-352-7671	Sequence 7671, Ap
27	174.5	8.6	358	US-09-328-352-5910	Sequence 5910, Ap

28	174.5	8.6	510	US-09-252-991A-25238	Sequence 25238, A
29	173.5	8.6	495	US-09-252-991A-23319	Sequence 23319, A
30	173	8.6	365	US-09-328-352-7412	Sequence 7412, Ap
31	171	8.5	386	US-09-328-352-4722	Sequence 4722, Ap
32	168	8.3	336	US-09-328-352-8202	Sequence 8202, Ap
33	164.5	8.1	314	US-09-252-991A-20336	Sequence 20336, A
34	161.5	8.0	357	US-09-252-991A-17970	Sequence 17970, A
35	161.5	8.0	399	US-09-252-991A-23589	Sequence 23589, A
36	161	8.0	373	US-09-252-991A-28902	Sequence 28902, A
37	157.5	7.8	374	US-09-328-352-5275	Sequence 5275, Ap
38	155.5	7.7	349	US-09-252-991A-28283	Sequence 28283, A
39	153	7.6	397	US-09-252-991A-23027	Sequence 23027, A
40	151.5	7.5	377	US-09-328-352-7154	Sequence 7154, Ap
41	148.5	7.4	420	US-09-252-991A-30038	Sequence 30038, A
42	148	7.3	411	US-09-328-352-6719	Sequence 6719, Ap
43	143	7.1	379	US-09-252-991A-26160	Sequence 26160, A
44	138.5	6.9	416	US-09-252-991A-19218	Sequence 19218, A
45	137	6.8	417	US-09-252-991A-28413	Sequence 28413, A

ALIGNMENTS

RESULT 1
US-09-328-352-4199
Sequence 4199, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Bretton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4199
LENGTH: 424
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4199

Query Match	39.7%	Score 802.5;	DB 4;	Length 424;
Best Local Similarity	43.7%	Pred. No. 2.2e-70;		
Matches 181;	Conservative 80;	Mismatches 136;	Indels 17;	Gaps 8;
QY	8	AMRAALAAVALVSSCGKGDAAOGGPRGRAPVGVVTHVHTVATLTTELPORTL	67	
DB	18	ALTRCALATSTIALV--GCSKSGDEKQAAAAQKMPAE--VGVIVAOPQSVQSVELSGRT	74	
QY	68	ESLRTADVRAQVGGIIQKRLFOEGSYRAGOPLYQIISSTYEANLESARAOATATIA	127	
DB	75	SAYISEVRQTSVILLKRLPABSGYVRREGALYELDSRTIRATLENNKSLLOOQANLA	134	
QY	128	KADADLARVPLVAEAVSRQEDAAVTAKRSAGVKAQAQAKTSAGINLRSRTAPI	187	
DB	135	SLRFLKRLRYQVLSNNAVSQKQEDYDILGQVNVAAQVAAKAQVTNNVLDGYSTISPI	194	
QY	188	SGRFGQKSVSEGTILNAGDTTVALITQTNPMYNNVQASAEVKKARQIAEGULAADG	247	
DB	195	SGQSGRSVAGALVTNNQDPLVTIOQLPIYVDINQSSLELRLRQOOLSKGSLNNSNN	254	
QY	248	VIAVGIFEDQGVYPEKGRLLFADPVVNESTGQITLRAAVPNDONILMPGLYVVLMDQV	307	
DB	255	T-KYKALIBDSYPIPIGQALFSPASVNOGTGTTLLAIVSNPHILLPGMYTTAOLVQG	313	
QY	308	AVDNAFVPPQOAVTR--GAKDTWIVVAQSGMEPREVTVAQOQGTNNIVTSLGDKQKV	365	
DB	314	VVPNAVYIIPQALITRLPTGQAVMLVNAKGVESRPVETSGVQGNQNNIVTNGLAADKVI	373	
QY	366	VEGISTIGITGAKVYTKEMAS--SENQAAAP-----QSGVQTASEAKTASEA	411	
DB	374	VVG--VAKVKEGQVSAKPYQAOAPANSQGAAPNAKAPQSG--KPQABQKASNA	424	

```
RESULT 2
US-09-252-991A-16988
; Sequence 16988, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16988
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16988

Query Match
Best Local Similarity 37.3%; Score 753; DB 4; Length 453;
Best Local Similarity 42.1%; Pred. No. 1.8e-65;
Matches 174; Conservative 68; Mismatches 125; Indels 46; Gaps 10;

QY 8 AMRAAALAAVAL-VLSSCGKGDAAOGGAPAGREAPAV---VGVTVHPQTALIVE 62
DB 76 AMRVLPVPLVLAISALSGGKS-----EAPPAQTPVEGIVTLEAQTTLATE 123
QY 63 LPRGLESRTADVRAOVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLATA 122
DB 124 LPRGNARFIAEVRQVNGIIILKRLFKESDVKAGQQLYQIDPATYEAQYQAQNLAST 183
QY 123 QATLAKADADLARVYPLVAEAESRQEDAAVTAKRSAGVKAQAALIKSAGINLNSR 182
DB 184 Q-----EQAGRYKLVAQVAVSKQY-----ADANAVALQSKAAVEGARIMRYTK 229
QY 183 ITAPISGFIGSKVSEGTILNAGDTTVALITROTNPVNVNTOSASEVWKLRQIAEGTL 242
DB 220 VLSPISGRIGRSATVEGLAVTNGQANAMATVQQLDEIYDVTPSTALRLRLRLASQQL 289
QY 243 -LAAQGVIAVGIKPDGTVYPEKGRLLFADPVVNESTGQITLRAVPNDONILMGLYVR 301
DB 290 EAGGNMAKAVSILKEDSGYPLEGRLEPSEVDESGTTRRAVPNNELLFGMPVH 349
QY 302 VLMDQVAVDNAAFVPPQAATRGAKD--TVMIYNAQSGMEPREVTVAQQQGTWIVTSGLK 359
DB 350 AQLQSGVAKKALILAPQGVTRDLKQATALVVAQMKVELRVTKADRVIGDKMLVTEGLN 409
QY 360 DDDKVVVVGISIA--GITCAKAVTPEKMASSENOAAPOSGVQTSSEATASE 410
DB 410 AGDKITTEGLQFVQPGVE-VKTVPAKNVASQAKADAAAP-----AKTDSK 452
```

```
US-09-252-991A-30590
; Sequence 30590, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30590

RESULT 3
US-09-252-991A-30590
; Sequence 30590, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30590

Query Match
Best Local Similarity 32.0%; Score 647; DB 4; Length 447;
Best Local Similarity 40.8%; Pred. No. 4.5e-55;
Matches 146; Conservative 66; Mismatches 136; Indels 10; Gaps 3;

QY 14 LAAAVATLVSSCGKGDAAOGGAPAGREAPAVVGVTVHPQTALVTELPGRLSRLTA 73
DB 60 LFLSLGILLQCC-----DSKEVQAQAPPAKVSLSIQOSVFNSENLPAKVAFRTA 112
QY 74 DVRAOVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLATQATLAKADL 133
DB 113 EIRPQVGGIIIEKVLFFKQSGSEVRAGQALYKINSETFEADVNSRSLNKAEEVRLKVQL 172
QY 134 ARYKPLVAEAESRQEDAAVTAKRSAGVKAQAALIKSAGINLNSRITAPISGFIG 193
DB 173 ERIEQLPSNAISKQEVNMAQAYRQALADVAQMKKALLARQNLQVYTVAPISGRIGQ 232
QY 194 SKVSEGTILNAGDTTVALITROTNPVNVNTOSASEVWKLRQIAEGTLAAGVIAVGI 253
```

```

Db 233 SFTVIGALVGGQDNTVMATIQIDKVVYDVKQSVSEYERLQALQSGE-LSANSDKTVRI 291
QY 254 KFDSTVPEKGRLLFADPVNVESTGQITLRAAVNDQNIIMPGLHYEVLMDQVADNAF 313
Db 292 TMSHQPINVTAKMLFEDINVDPEGDVTFRIEVNTERKLLPGMYRVNIDRASIPDL 351
QY 314 VVPOQAVTR--GAKDTWMI VNAQGMPEPRTVAQOQGTNMI VTSGLKDGKRVVEGI 369
Db 352 LVPAQAIQRNIGSGEPQVYVINAQGTAEIRPIEIGQYQGFYANGLKGVKRVVEGI 409

```

RESULT 5
US-09-252-991A-27975
Sequence 27975, Application US/09252991A
Patent No. 6551795

```

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27975
LENGTH: 414
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27975

```

Query Match
Best Local Similarity 31.8%; Score 642; DB 4; Length 414;
Matches 157; Conservative 65; Mismatches 157; Indels 22; Gaps 5;

```

QY 14 LAAAVALTSSCGKGDAAGGQAPAGREAPAPV-VGVTVHPQTVALTVLPGRLLESURT 72
Db 31 LAAVALFLGCEEAAD-----AGKTAEPAEVGVVAPPAIGITSELPGLLEARTQ 83
QY 73 ADVAQVGGIIQKRLFOGGSYVRAGQPLVYQIDSSSTEYANLESARAOATLAKADAD 132
Db 84 AEVARAVAGIVTRLYEEGQDVPRAGTVLFQIDPAFLKALDISGALARAASHAAADK 143
QY 133 LARKPLVAAAVSROEYDAAVTAKSAEAGVKAQAQAIKSGINLNSRITAPISGFIG 192
Db 144 LKRYADILKRAISEREYTEQOTARQALQIASAKLEQARLRIGAVTATIDGRAR 203
QY 193 QSKVSEGLTLAGDPTVATIRGTNPVYVNTQSSASEVMKLRQIAESKLL-AADGVYAV 251
Db 204 RALVTGELVEDEPTPLIRVEQIDPIVNSQPAEVAAMQRIREQVGVADKDIATV 263
QY 252 GIKFDDGTVYEKRLLEFADPVNVESTGQITLRAAVNDQNIIMPGLHYEVLMDQVADN 311
Db 264 RLVLADGSEYLAGELLFSDLAIVPGTDTIMRLFPNPHRELLPGGVYVRLGRAVNPQ 323
QY 312 AFVVPQAVTRGADVTM-VNAQGMPEPRTVAQOQGTNMI VTSGLKDGKRVVEGIS 370
Db 324 AITVPDALIRTAQSAVAVKVNPGQLVEDVEVRADTLQGRWITISSGLKGERIVENAA 383
QY 371 IAGITGAKVTTPKEMASSENOAAPQSGVQTAASEAKTASEA 411
Db 384 -----QTAAGSSVQAVVRQPASADAPSPPLAASPA 412

```

RESULT 6
US-09-328-352-4570
Sequence 4570, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4570
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4570

```

Query Match
Best Local Similarity 31.6%; Score 451.5; DB 4; Length 409;
Matches 131; Conservative 82; Mismatches 180; Indels 21; Gaps 7;

```

QY 4 YAFKMAAALAAVALVLSGCGKGDAAGGQAPAGREAPAPVGVTVHPQTVALTVL 63
Db 11 FALSAIFVALLATGSEFML-LHENADAKAPTAQQAATVDAPVVS--KITDMQEV 65
QY 64 PGRLESURTADVRAQVGGIIQKRLFOGGSYVRAGQPLVYQIDSSSTEYANLESARAOATL 123
Db 66 SGRLEALDQVDIRQVSGKLIAMHFKDGSIVKKGDLFTIDPRFEALNRKAQOLASAE 125
QY 124 ATAKADADLARYKPLVAAEAVSROEYDAAVTAKSAEAGVKAQAQAIKSGINLNSR 183
Db 126 AQVTYTSNLSRIQRLQSNVSVSDELLENDRASANMLQARAAVQSRRLLETRI 185
QY 184 TAPISGFIGQSKVSEGLTLAGD-TTVLATIRGTNPVYVNTQSSASEVMKLR--RQIAE 239
Db 186 TAPVSGRISRAEAVVGVNVAAGNAQVLTISVSRLYASFVDEQTYLKYISNQRSQAQ 245
QY 240 GKLLAADGVIAVGIKFDGTVYPERKRLFPADPVNVESTGQITLRAAVNDQNIIMPGL 299
Db 246 -----VPTWGLMETGFTREGTINSIDNNLTISGITRRAATFDPNGLPELY 296
QY 300 VRVLMDQAVADNAFVVPQAV-TRGAKDTWMI VNAQGMPEPRTVAQOQGTNMI VTSGL 358
Db 297 ARIRLGGQPRPALLISFTAVGVQDQKRFVYVVDAAKQTVAREKLGAAQDGLQIVNSGL 356
QY 359 KDGKRVVEGISIAGITGAKVTTPKEMASSENOAAPQSGVQTAASEAKTASEA 412
Db 357 QAGDRIVNGLQ--RIKPGDPTVPHLVPWMSQITATSPQPPQPTDKTSTPAK 408

```

RESULT 7
US-09-252-991A-31636
Sequence 31636, Application US/09252991A
Patent No. 6551795

```

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 31636
LENGTH: 455
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31636

```

Query Match
Best Local Similarity 32.7%; Score 448; DB 4; Length 455;
Matches 133; Conservative 75; Mismatches 177; Indels 22; Gaps 9;

```

QY 14 LAAAVALTSSCGKGDAAGGQAPAGREAPAPVGVTVHPQTVALTVLPGRLLESURT 73

```

```

Db      58 LALAALVLSACGKAPETTOG-----MAAPKSVAAEVIQPLNEMWDEFTGRLEAPSPV 110
Qy      74 DYRAVGGIQLKRLFOEGSVYRAGPLVQIDSTYEANLESRAQLATQAATLAAADL 133
Db      111 ELRPVSGTIDVAHEGALVKKGDLLEQIDPREPEAEVKLEAOVQQAQAQASVVEA 170
Qy      134 ARYKPLVAAEAVSROEYDAAVTAKRSAAEGVAAQAATKASAGININRSRITAPISGFIQ 193
Db      171 QRGELRLASMAISAFAPARTTAQAEAKAANAATQAQDAARLANSFTRITPIDGRYSR 230
Qy      194 SKVSEGTLLMAGDTTLATIRQTNPMYVNTQASAEVWK--LRQIAEGKLLAADGVIA 250
Db      221 AEVTAGNIVNSGE-TLLTTLVSTDVYVAFDADERVFLKVELAQ--AGRDTRESEPV 287
Qy      251 VGIRKDDGTVPYPERKLLFADPVNVESTQITLRAAVPNQDQILMPGLVRYRLMQVAVD 310
Db      288 LGLSESDGN--PHLGRLDLDVQVNPRTGTLGRVAFDPAKEFPGLVRLKLGSKTY 345
Qy      311 NAFVVPQOAV-TRGAKDTVMIVNAAGMEPREVTAAQOQGTNMIYTSGLKDGKTYVEGI 369
Db      346 AATILKDEAVGTDLGKKFVLVDGDKNTYRTVEVMPKLEGLRIVRSGLSKGDRIVAVGL 405
Qy      370 SIAGITGAKKVTYK--EWASSENQA--AAPQSGVQTAASEAKTASAE 412
Db      406 Q--RVRPGQVDPQKVEWASADTLATLARLRQSVGDSPEPKVAASKD 450

```

RESULT 8

```

US-09-252-991A-31241
; Sequence 31241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31241
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31241

```

```

Query Match      21.2% Score 429; DB 4; Length 429;
Best Local Similarity 31.0%; Pred. No. 1,1e-33;
Matches 119; Conservative 59; Mismatches 172; Indels 34; Gaps 7;

```

```

Qy      6 FKAMR-----AAALAAVALVL-----SSCGKGDPAAGGQOPAGR--EAPAVVGVTV 52
Db      13 FRTLARPLWLTALFAAIVGLVWMLAAPASAPSDRPGKGRGALPRLANLTVGVARV 72
Qy      53 HPQVALVELPGRLESLELTADVRAQVGGIIQKRLFOEGSVYRAGPLVQIDSTYEANL 112
Db      73 ECGDIALHFNALGTVAFTVAVKPRVNGELVKVLFQEGQEVKAGDILLAVDPRTYKAL 132
Qy      113 ESARQQLATQAQTLTAKADADLARYPVIAEAESVROEYDAANTARSAEAGKAAQAIAIX 172
Db      133 AQAESTLQONQALKNALETDLQRYGLVAEDSIKQTLDTQEAQVRLQGLRTTQGOVD 192
Qy      173 SAGININRSRITAPISGFIQSGVSEGTLLNAGDTTLATIRQTNPMYVNTQASAEVWK 232
Db      193 DARLALTFEVAAPISGRIGLRQVNDIGNLVTSQDTTLPVLTQVPERISVFSLEPQQIGT 252
Qy      233 LRRQI-ABGKLLA-----DGVIAVGKFDGTVYPERKLLFADPVNVESTQITLRA 285
Db      253 VVEQNNNGPKLAVTLALDRNQDKVLA-----EGTLTTLDNQIDTDTTGVKKA 299

```

```

Qy      286 AVPNQDQILMPGLVRYRLMQDAVNAFVVPQAVTRGAKDT-VMIVNAAGMEPREVTV 344
Db      300 RFENADGKLFNQGVNRRILAHTLKGLVLTIPANAVQGTNGIYVVVGADKVKSGRSVAI 359
Qy      345 AQQOQTNMIYTSGLKDGKTYVEG 368
Db      360 GTSNERVYVESGLKAGEQVVEG 383

```

RESULT 9

```

US-09-252-991A-24388
; Sequence 24388, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24388
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24388

```

```

Query Match      19.5% Score 393; DB 4; Length 502;
Best Local Similarity 33.7%; Pred. No. 4.9e-30;
Matches 118; Conservative 50; Mismatches 162; Indels 20; Gaps 5;

```

```

Qy      28 GDAAGGCGPAGRAPAPV--VGVTVHPQTVATLYELPGRLESLELTADVRAQVGGIIQ 84
Db      141 GGMAMLGSDAGAKAPAPAPVAVVAVRERDVQVSGIGTVSLHNVVIRTQIDGQLT 200
Qy      85 KRLFOEGSVYRAGPLVQIDSTYEANLESRAQLATQAATLAAADADLARYKPLVAAEA 144
Db      201 RLIVSEGGWVAGELATIDRAVVAALQQAASRASQAOLKSAEODLQHRSLYARARA 260
Qy      145 VSRQDYDAAVTAKRSAAEGVAAQAATKASAGININRSRITAPISGFIQSGVSEGTLLNA 204
Db      261 VSRQLDDQQAATVDQLRATTLKANDATTNAEVRRLSTRTITSPVSGKGIIRVVDGNLVRY 320
Qy      205 GDTTVLATIRQTNPMYVNTQASAEVWKLRQI--AEGKLLAADGVIAVGIRKFDGTVY 261
Db      321 GDSIGLFVETQIAPISVFSIQEQQLPOLQALGGAAVRAVSRDGSALG----- 371
Qy      262 PEKGRLLFADPVNVESTQITLRAAVPNQDQILMPGLVRYRLMQVAVDNAFVVPQAVT 321
Db      372 --EGRLTIDNQIDSTQITIRVRSFDRQARLMPGQFVAVSLHGVARRDLVSSKXVR 429
Qy      322 RGAKDTVMIVNAAGMEPREVTAAQ--QGTNMIYTSGLKDGKTYVEGIS 370
Db      430 RGLSENFYRYAADRVAEAVPRVLQDIDGLS--VVEGLASGDQVVDGHS 477

```

RESULT 10

```

US-09-199-637A-149
; Sequence 149, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Lawrence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Cao, Hui
; APPLICANT: Tan, Man-Wah
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John

```

```

; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: SEQUENCES AND USES THEREOF
; CURRENT APPLICATION NUMBER: 00786/361002
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 149
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-149

Query Match      18.4%; Score 370.5; DB 4; Length 202;
Best Local Similarity 37.9%; Pred. No. 1.9e-28;
Matches 80; Conservative 43; Mismatches 73; Indels 15; Gaps 5;

OY 205 GDTVLATIRQTNPMYVAVTOSASEVMKLRQIAEGKL-LADGVIAVGIRKPDGTVPY 263
DB 1 GOANMATVQQLDPIYVDTQSTALIRMRRELASGQLEBAGDNAAKVSLEKLEDSQYPL 60
OY 264 KGRLLFADPVPVNESGQITLRAAVNDQNIIMPGLYRVLDQVAVDVAFFVPOQAVTRG 323
DB 61 EGRLEFSEVSDGEGSTIRAVFPNPNNELPGFVFVAHQLOEGVKOKAIIAPQOGVTRD 120
OY 324 AKD--TWIIVNAQGMREPREVVAQOQGTNMTVTSGLKDGKVVVEGSI--GITGAKK 379
DB 121 LKGAQTAIVVNAQNKVRELRAVIGADKVIQKMLVTGLNAGDKITTEGLOFQVGVE-VKT 179
OY 380 VTPKEMASSENQAAAPQSGVQTAASEAKTASE 410
DB 180 VPAKVVASAQKADAP-----AKTDSK 201

RESULT 11
US-09-252-991A-17691
; Sequence 17691, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17691
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-17691

Query Match      17.6%; Score 355; DB 4; Length 458;
Best Local Similarity 27.6%; Pred. No. 2.3e-26;
Matches 102; Conservative 78; Mismatches 169; Indels 20; Gaps 7;

OY 18 VALVSSCGKGDAQGGOPAGREAPRVGVVTHPQTVALITVLPGRLESITADVRA 77
DB 88 VAALIAAGGESA-----PPGAASAPPSVPAEVAEVPVTPVPAEFTGSLTAIVEQVELRP 140
OY 78 QVGGIIQKRLFOEGSYVYAGQPLVQIDSTYEANLESRAQOLATAQATLAKADLARYK 137
DB 141 RVAGYIGQVTVPEGRVLEKGGQQLFLIDRVFKKAQDABARARLREBAALALAFHEHBAE 200
OY 138 PLVAAEAVSRQETDAVTAKSAEAGVAAQAAIKSAGININRSRTITAPISGFIQSKYS 197
DB 201 LLVARKVAVRERLDSATASRNASKAQVDAAABALDAQQLDGLGFTRTVAPIGRGVGHIOVT 260
```

```

OY 198 EGTLLMAGDTVLATIRQTNPMYV--NVTOSASEVMKLRQIAEGKLAAADGVIAVGIRK 255
DB 261 EGNVTVNG-VTALNISIVSDPLYVYFDVDERTYIQLALAPTRGRSEEQAPR---KVALLT 316
OY 256 DDGTVYPEKGRLLPADPVNVESTGQITLRAAVNDQNIIMPGLYRVLDQVAVDVAFFV 315
DB 317 DES--YGRSSRLDPLANNAADRGTGTVRAVAVNDPDGQITPGLPAKVFLETGKPPRAQVIV 374
OY 316 PQQAV-TRGAKDTMIYVNAQGMREPREVVAQOQGTNMTVTSGLKDGKVVVEGSIANGI 374
DB 375 ADHSITGDQGRRYLVVDENKNTQYRPVELSPMVDGLRVAVQGLQPGERIVVKG---V 430
OY 375 TGAKKVTPK 383
DB 431 RPDQITPR 439

RESULT 12
US-09-252-991A-24485
; Sequence 24485, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24485
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24485

Query Match      13.7%; Score 277; DB 4; Length 394;
Best Local Similarity 27.1%; Pred. No. 8.4e-19;
Matches 104; Conservative 71; Mismatches 141; Indels 68; Gaps 14;

OY 14 LAAVALVLSGCG-KGDDAAGGOPAGREAPRVGVVTHPQTVALITVLP----- 64
DB 28 LAAVIAVVAIIAGIKVYSIRQ--QIALPSAKPEPTSV-----TASIAEKRPQOSRLPAI 79
OY 65 GRLESRTADVRAQVGGIIQKRLFOEGSYVYAGQPLVQIDSTYEANLESRAQOLATAQA 124
DB 80 GSKAPQGVITLTAESGTVADVFLPSGDQYKLDQPLQLDSVDVEATLR-----TREA 132
OY 125 TLAAADLARYKPLVAEAVSRQETD--AAVTAKSAEAGVAAQAAIKSAGININRSR 182
DB 133 DLGIARAEOGRRELISKAISKSEFDRLLAAQMAKTSATV-----AELKAA---LAKR 183
OY 183 ITAPISGFIQSKYSEGTLLMAGDTVLATIRQTNPMYVAVTOSASEVMKLR-QIAEGK 241
DB 184 VLAFAAGTIGIRQVDVDVSPG--TPIATLQDLSILLDFHLPEDQFPPLSRQOLVKVR 241
OY 242 LLAADGVIAVGIRKPDGTVPYPEKGRLLPADPVNVESTGQITLRAAVNDQNIIMPGLYV- 300
DB 242 VAAVPGQV-----FD-----AETAINPKDNETRNIQVRALENPQKLLPGMFAN 288
OY 301 -----RVLMDOQAV-----DNAFVVPQQAVTRGAKDTMIYVNAQGMREPREVTV 344
DB 289 LEVWLPGEORVVVVPETAIFETLYGDSIYVVGQKDEQGVSKDKGQPOOVVRRFRVRI 348
OY 345 AQQOQGTWITYTSGLXKDGKVVVEG 368
DB 349 GERREGIAVVLLEGGEQVVTSG 372
```

```
RESULT 13
US-09-328-7897
; Sequence 7897, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7897
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7897

Query Match      13.5%; Score 272.5; DB 4; Length 452;
Best Local Similarity 23.1%; Pred. No. 2,9e-18;
Matches 99; Conservative 62; Mismatches 127; Indels 141; Gaps 11;

QY 65 GRLESLRTADVRAQVGGIIQKRLFOEGSYVRAGQPLVYDIDSTYE-----ANLESARA 117
DB 60 GTLDATKLISVGAQVSGVKKMYVQLGDQVKGQOLIAQIDSTTOENSLKTSIDANIKMSEA 119
QY 118 QLATQAATLAKADADLARYKPLVAABEAVSROEYDAAVTAKRGAAGVRAQAQAIKISAGI- 176
DB 120 QRLQAIASINEKQLEFRQOQMYAQDAPTRADLESABAAYKTAQAQVAKLDAQISAKIT 179
QY 177 -----NINRSRTAPISGFIGOSKVSSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEV 230
DB 180 RSTAGTNGYTRIVAPTGTVAIVTEBQTVNA-----NQSAPTVIKATLQI-- 228
QY 221 MTLRQRIABGKLADGVIAVGIRKDDGTIVPEKRLPADPVNVESTGQITLRAAVEND 290
DB 229 MTKAQVSEADIMKV-----EKQOQVYFTLLGDETRKRYATLRQIEPAP 271
QY 291 QNI-----LMPGLYVRVLM-----QVAVD-----NAFVVPQAV 320
DB 272 DSISESNSTSTSTSSAVYNAALPDVPTDGLKIDMTAQVYIYVNSAKNALVPSNHL 331
QY 331 T-----RGAK-----DTVM 329
DB 332 SSKQPSGGRKQGSADKASSTPSABERKHQNGARLERLNLTPKQKLIHQKATLSYVR 391
QY 330 IYNAQSGMEPREVYTAQOQGTWVIYTSGLKDGKVVYEGISIAGITGAKKTPKEMASSE 389
DB 332 VLAQAGTTPQIILVGINNRVNAQVLAKIKGDQVVIAD-----SSE 433
QY 390 NOAAPQSG 398
DB 434 NSASANSNG 442

RESULT 14
US-09-252-991A-20610
; Sequence 20610, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20610
; LENGTH: 399
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20610

Query Match      13.2%; Score 267; DB 4; Length 399;
Best Local Similarity 25.6%; Pred. No. 8.2e-18;
Matches 93; Conservative 64; Mismatches 170; Indels 36; Gaps 8;

QY 11 AAALAAVALVSSCGKGDAAQGGPAGRAPAPVGVTVHPQVATLTVELPGRLESL 70
DB 38 SALLCAAVIGIAYVYATGSAKKDAGFRAG--YYPVKYALASVERVPRVVDGCGELEAG 95
QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGQPLVYDIDSTYEANLESARAQATLAAQAD 130
DB 96 ROVVAABEAGRIIRIAPFESGQVQOQGLVLQVNDVAFQELIRLKLQLRNABILHARA- 154
QY 131 ADLARYKPLVAABEAVSROEYDAAVTAKRGAAGVRAQAQAIKISAGININRSRTAPISGF 190
DB 155 -----RRLVERNVASOEQLDNAAVARDMALGAARQTOALI-----DQAIRAPFSGQ 201
QY 191 IGQSVSEGTLLNAGDTTVLATIRQTNPMYVNTQSASEVVKLRBQRIABGKLADGVIA 250
DB 202 LGIRRVHLGQYLVGAEPVASLVDAKRTLSNFSLESTSPKLGQP-----LE 249
QY 251 VQIKFDDGTIVPEKGRLLPADPVNVESTGQITLRAAVPNDQNIIMPGLYVRVLMDOYAVD 310
DB 250 VLVDAYPGRSFP--ARISAIIDPLIGKSR-TVQVQALLDNPGLLAAGMFASIRSRKADA 306
QY 311 NAFVVPQAVTRGA-KQTVMLVNAQSGMEPREVYTAQOQGTW-----IVTSGLDQDGKVV 365
DB 307 PSLSVPEYTAHYTVAGTVFVAHQDGRPLSAKRVSIRIGRMQGRVEILLQGLAEGDRVV 366
QY 366 VEG 368
DB 367 TSG 369

RESULT 15
US-09-252-991A-28543
; Sequence 28543, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28543
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28543

Query Match      12.7%; Score 256; DB 4; Length 406;
Best Local Similarity 26.9%; Pred. No. 1e-16;
Matches 100; Conservative 68; Mismatches 154; Indels 50; Gaps 11;

QY 10 RAALAAVALVSSCGKGDAAQGGPAGRAPAPVGVTVHPQV-----ALTVELP 64
DB 39 RVALPALICAGLLVCGG-----APPPEEHVRAVLAQVTKMAEFASATST 83
QY 65 GRLESLRTADVRAQVGGIIQKRLFOEGSYVRAGQPLVYDIDSTYEANLESARAQATLAAQ 124
DB 84 GDQARVADQSPFVGGKIVYRLVDVGHVAAQVTLARLDPQDQSRNVENQAQVAAQQA 143
QY 125 TLAAADADLARYKPLVAABEAVSROEYDAAVTAKRGAAGVRAQAQAIKISAGININRSRT 184
```


Db 144 QSKLADNYQROKALLPKYTSQSEYDQALASVRSQSSLKAAQOQANAPDLISYTEIR 203
Qy 185 ADISGEFIGSKVSEGLTNAGDTYVATIRQ--TNPVNVNTQS--ASEVMKLRQIAEG 240-
Db 204 ASDAGVITARQAEVQVQA--TVPIFLTARDGERDAVNVYESLFSHDVDQRLTVS-- 259
Qy 241 KLLADGVIAVIRKDDGVIVPEKGRLLFADPVNVNESTGQITLRA--AVPNDQITLMPG 297
Db 260 -LLGKREYVA-----SGKVRITPTVDERSGTTLKVKGLDVSFPAEWSL---G 302
Qy 298 LVYRVLMDQVAVDNAFVVPQQAIVTR-GAKDTVMITNAGSGMREYTVAAQOQGTWITYS 356
Db 303 SVVNASVAAPA-EHSVTVLFWMSLSKVBQPAVWLIDQOQKARLPQVRAVRAKSEKVIDG 361
Qy 357 GLKDGDKVVEG 368
Db 362 GLEAGQTVTVTG 373

RESULT 16
US-09-252-991A-21202
; Sequence 21202, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21202
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21202

Query Match 12.2%; Score 245.5; DB 4; Length 389;
Best Local Similarity 22.0%; Pred. No. 1e-15;
Matches 90; Conservative 72; Mismatches 189; Indels 59; Gaps 8;
Qy 4 YAFKMRRAALAAVALVLSGCGKGDAAOGQAPAGRAPAPVGVTVVHPQTV-A-LTVE 62
Db 22 HMYRITPLVALSLFESSLFLAAGNGTPPPAARPA-----IVQOPAGEVSQA 70
Qy 63 LFGRLSLEKRTADVBAOVGGIIOKRLFQEGSVYRAGOPLYQIDSSYTEANLESARAQLATA 122
Db 71 FFGERRAHHEBELARIGSKYIRLVEVGERVKDQPLAELDPQDVRLQLEAARQVSA 130
Qy 123 QATLAKADADLARYKPLVAEAVSRQEYDAAVTAKRSAGVKAQAQAIKSGAGINLNSR 182
Db 131 EANIQTVAEYRRTLLDRMLVSHSGFENIONSVRAGEARLKQIRAFENVADNOAGYAV 190
Qy 183 ITAPISGTIGSKVSEGLTNAGDTYVATIRQTNPMVNVNTQASFEVMKLRQIAEKL 242
Db 191 LRSPDQGVIAARRVEVQVAAQGVFSLADGEREVLIGLPEHSFERFRIGQPVSELM 250
Qy 243 LAAGVIAVGIKFPDGTIVPEK---GRLFPADVNESTGQ-----ITLAAVNDQNI 293
Db 251 SQDRRRFAGHIRELSPADPOSRTFAARVAFDRAITPRLQOSARVYVAAAEVAP----- 305
Qy 294 LMPGLYVRLMDQVAVDNAFVVPQQAIVTRGAKDTVMITNAGSGMREYTVAAQO----- 347
Db 306 -----LSVPLSALTAEAGQAFVWV-----VEPGSSTRQAVRTGP 341
Qy 348 -QGTNMTVTSGLKXGDKVNVGIGISAGITGAKKYTPKEMASSENQAAAPQ 396
Db 342 YAEGRVVPVLEGLKAGDWVAVATGVQV--LREGQVRFIDRARTVYLAKE 389

RESULT 17
US-09-252-991A-28510
; Sequence 28510, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28510
; LENGTH: 770
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28510

Query Match 11.8%; Score 237.5; DB 4; Length 770;
Best Local Similarity 24.6%; Pred. No. 1.8e-14;
Matches 92; Conservative 61; Mismatches 176; Indels 45; Gaps 10;
Qy 8 AMRAALAA--AVLVLSGCGKGDAAOGQAPAGRAPAPVGVTVVHPQTV-----AL 59
Db 414 AMKFPSLAGLGFALLISGCG-----DEPPAPRPVLYTVVTKLKNDDIGRF 461
Qy 60 TVEPLGRLSLEKRTADVBAOVGGIIOKRLFQEGSVYRAGOPLYQIDSSYTEANLESARAQL 119
Db 462 AGSIQAYVESV---LGRFTNGRIASRLFDVGDVGVGALLATLDPDQOQQLAASQDL 517
Qy 120 ATAQATLAKADADLARYKPLVAEAVSRQEYDAAVTAKRSAGVKAQAQAIKSGAGINLN 179
Db 518 ASAEQLIDAQANARROEELFARSVTAQARLDARTLKTQSAQSFQDAKAIVQOARDLS 577
Qy 180 RSRITAPISGTIGSKVSEGLTNAGDTYVATIRQTNPMVNVNTQASFEVMKLRQIAE 239
Db 578 YTRLVTDFDGVITTHMAEQVVSAGO---AVVTLARP-----EVREAFFDLPTVEAE 627
Qy 240 GKLLAADGVIAVGIKFPDGTIVPEK---KGRLLFADVNESTGQITLRAAVPNDQNIIMP 296
Db 628 S--LPADARPLVSAQLD-----PQARTGSIRELGPQADASTRRRRLSLAQTPPEARL 680
Qy 297 GLYVRLMDQVAVDNAFVVPQQAIVTRGAKDTVMIVNA-QGMEPREYTVAAQOQGTNMT 355
Db 681 GSTIQVQLSSAGSVSVLPASVLLERDKTQVWVVDKQSSVALREVQVLSRDERQVIG 740
Qy 356 SGLKDGDKVVEGI 369
Db 741 QGLADGDRVRVAGV 754
RESULT 18
US-09-328-352-8183
; Sequence 8183, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8183
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8183

```

Query Match 10.9%; Score 220; DB 4; Length 500;
Best Local Similarity 22.7%; Pred. No. 4,9e-11;
Matches 72; Conservative 57; Mismatches 112; Indels 76; Gaps 10,

OY 71 RTADYRAQVGGIIQKRLFOE--GSYVRAGOPLYQIDISTYEANLESARAQATATLAK 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 132 QVAILQTRASGFVQ-RVYGHAVGDMVTQGSPIADISIPEMTGEOTEFIALVTRGDRSL- 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 129 ADADLARYPELVAAEAVEROEYDAATYAKRSAGGVAAQAATISAGINLN----- 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 189 -----10ASRQRQLLGIPOVINQVERTH 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 180 ----RSRTAPISGFIGOSKVEGTLNAGDTVIATPQTNPMYVNTQGSSEWMLKR 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 214 RVQSNMTLSAPVSGFIDSLSEVRNGMALAMQOT-LVTIKGISPLWEASVEAOIATIKR 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 236 QIAGKTLAADGVIAVGIKEDDGT-VPE--KGRLLFADPVNVESTQITLRAAVPNDON 292
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 272 -----GMKEATEFVAVPQIVSGKVIDILPLIDSTSRITKRIELPREG 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 293 ILMPLGYRVYLVADQVAVDNDAFVVPQQAATV-TAKDITMIVNAQSGMEPRETVAAQOGSTN 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 316 MKPQGFAPVAKESNNPQSN-LVIPQAVIRTRGNVAVIGHQGRFEPVAVQLQSDGSK 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 352 WIVTSGLKDGDKKVVEG 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 375 IALQLGLKAGQKVVISG 391

RESULT 19
US-09-328-352-5003
Sequence 5003, Application US/09328352
Patent No. 6552958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5003
LENGTH: 367
TYPE: PR
ORGANISM: Acinetobacter baumannii
US-09-328-352-5003

Query Match 10.7%; Score 216.5; DB 4; Length 367;
Best Local Similarity 24.7%; Pred. No. 6,6e-13;
Matches 93; Conservative 63; Mismatches 165; Indels 55; Gaps 14;

OY 40 RAP-----APVGV-----VTVHPQVATLVLEPGRLESITRADYRAQVGGIIQKRLFOEG 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 25 KEAPKTEERPYVWVTPSTLHEOK-----SYAGDVQARQOTALAFRGGQVTRARYVDVG 79
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 92 SYVRAGOPLYQIDISTYEANLESARAQATATLAKADADLARYKPLVAAEAVSRQED 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80 DRYRQGVQVLAKLDEVDQAQLQNLNAKAQQLDNAQAQAAKTAASBELKFKQQLPINAVERSCYD 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 152 AAVTAKRSAGVKAQAQAAIKSAGINLN--SRITAPISGFIGOSKVEGTLNAGDTT 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 ---TYNQVDAAQAALQOARSNYEVSANQGVNQLVSNKNGVITFARNIEIGQVVAAGQAA 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 209 VLAITRQTNPMYVNTQGSSEWMLKRRLQIAEGKTLAADGVIAVGIKEDDGTVEP----- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 197 YQALADGEREVAVIGVEQA-----VSEIKGQAAMI-----TLMKSPNERF 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 264 KGRLLFADPVNVESTQITLRAAVPNDONILMPLGYRVYLVADQVAVANAFVVPQQAATVRG 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 AGYVAEVSAPADQSR-TFTYKVALIKGQSAIQLGQSKARVFFSSQTQ-NMSVPLSSVS-- 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 324 AKDT---VMI VNAQSGMEPRETVAAQOGTNMTVTSGLKDGDKKVVEGISTAGITGAKV 380

```

```

Db      294 ATDQPYVWVNAQTLRKPYTITGAGRDSVPVLSGLTNDWVVGIVHL--LRDKQKI 351
QY      381 TPKEWASSENOAAPQ 396
Db      352 HP---IDRENRAVKIQ 364

RESULT 20
US-09-252-991A-25447
; Sequence 25447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25447
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25447

Query Match      10.6%; Score 215; DB 4; Length 487;
Best Local Similarity 25.0%; Pred. No. 1,4e-12;
Matches 89; Conservative 60; Mismatches 157; Indels 50; Gaps 10;

QY      50 VTVAHOTVATLVELPGRLESIRTDVAVAGYGIIOKRLFOEGSVVRAGCPLOYIDSSTYE 109
      135 VSVRGSTIESSVSLGTLGQRRRYVDVAQASGQIRKIHVAGDVTBEGQLVBIPESTQ 194
      110 ANLESAR-----AQLATAQATLAKADADLARYKPLVAAEAVSROEYDAA---VTAK 157
      195 AKVDAGRSIEMLKAGLAEGRAQCYTLARQGYORQORLAAGARTTEDVSAQAQMLATQA 254
      158 R--SAEGVYRAQAQAISAGINLNRKITAPIGFGIOSKVSSEGLTUNA--GDPTVLATIR 214
      255 RIENYQAOIROAQASLSRSDENELGYTHIVAPMSGTVAVADAREGQTLNAQQOTPLILRIA 314
      215 QTNPMYVNVVQASAEWKKLRROI--AEGKLLAADG-----VIAVGIR---PDDGTV 260
      315 KLSPIYVAOVSEADIGRVKPMFAYTTTISGEGRRRTGKVRQILPPPPRLDOSNCGG 374
      261 YPEKGRLLFADPVVNESTGQ-----ITLRAVPNDQNILMPGLYVRVLMDOAVDNA 312
      375 SPTSG-----SGGSGSGRVVLYTVLVVDVDGDHQLMAEMTAQVFFVAATRENI 423
      313 FVDPQAVTRGAK--DPIVIVYNAQGMPEPEVYVAQQGCTNMIVYTSLKAKGDKVYV 366
      424 LTAIVAAIHDDGKGQVAVWVGSNGKQPSQRIRTGISDRLRVYVLAGLEGRDILM 479

RESULT 21
US-09-252-991A-19591
; Sequence 19591, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 19591
 ; LENGTH: 306
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-19591

Query Match	9.8%	Score 197.5;	DB 4;	Length 306;
Best Local Similarity	28.2%;	Pred. No. 3.6e-11;		
Matches	89;	Conservative	43;	Mismatches 121;
			Indels	63;
			Gaps	14;

```

QY      52  VHQPTVALIYELBGRLESRTADVRAOVGIIIOKLFQBGSVYRAGCPLOYIDISTYEAN 1111
Dh      45  VSAQVIRIAEVSQSVAEVAFVAD-----NQRVARGDPLVYIDRSTRILA 88
QY      112  LESARQAQLATQAOTTLAKADADLARYKPELVAAEAVSRQEYDAVTAQKSAEKGVAQAQAI 1717
Dh      89  VROREAELEQARETELFKORDEQWRMRQOL--AGAVSREEVANAAGRLLRIARARAAQAAL 146
QY      172  KSAGININRSRIAPISCFIGOSKSEGTLLNAGCTVLATIRQTNPMYVNAVTOASAEVM 2311
Dh      147  DQARLDIRRELSNPVDYVQTQARVQPGDYAAAGNTNFIYDRRS--FWVT--GYFEET 201
QY      232  KLR--QT--AEGKLLAADGVAVGKIFPDGYTYEKGKRLFPADPVNVESTGQITLPAV 2877
Dh      202  KLRNVQVQADATKIKMGFPLL-----DGRV-ASIRGV-AD--LMSPADSGLOVS 250
QY      288  PNDONI-LMEGLYVRLMDQVAVDNAFVPPQAVTRGAKDPTMIVNAGMGEPREVTVAQ 346
Dh      251  PNRSMWIRLAQRVPVRIEIDRVPA-----GVTLAAGMTGSVE-----VL 2888
QY      347  QOQT--NMIVTSGLCD 360
Dh      289  QAGARHRWRLTRWLOE 304

```

```

RESULT 22
US-09-328-352-7523
; Sequence 7523, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7523
; LENGTH: 367
; TYPE: prt
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7523

```

Query Match	9.1%;	Score 184.5;	DB 4;	Length 367;
Best Local Similarity	25.3%;	Pred. No. 9.2e-10;		
Matches 85;	Conservative 46;	Mismatches 114;	Indels 91;	Gaps 13;

```

QY 49 VVTAVHQYALVAVELFGLESJRTLDVRA-----QVGIIQKRLFOEGSVYRAGOR 99
      : : : : : : : : : : : : : : : : : :
Db 40 IVALSIIIMILKVIPLPSV--VKTDARVDVESTIAPKVSNGIIEIYIKDHQVKKGOL 97
      : : : : : : : : : : : : : : : : : :
QY 100 LVYIDSBSTEANLESRAQALATPAQATLAKA-----DADL----- 133
      : : : : : : : : : : : : : : : : : :
Db 98 LARIDARDYQALAEESNVYAKAQADLINEAMLAVERPVIREFEAOJLRKVEAGIKLTKO 157
      : : : : : : : : : : : : : : : : : :
QY 134 --ARYEPVAAEAVSR-----QEDAAVTAKRSA 160
      : : : : : : : : : : : : : : : : : :
Db 158 NTRARYEOLALGEBESRLITQOSKTYLTROYADLDSKEKVIADAYOLNOYKGIQVQAK--- 214
      : : : : : : : : : : : : : : : : : :
QY 161 EAQVRAAKAAIISAGINLNSRITAPISGFTGQSVNSGTLINNGDPTVLATIRQTPMY 220
      : : : : : : : : : : : : : : : : : :

```

Db 215 QAAKQQAALDPAKLTANSTYETIRAPIDMGIGKSAVNPFAGANP--LMMVVPIDQY 272

QY 221 VNVVQSAEWMYKLRQIAEGKTLADGVIAVIGIKPDGTVYPEKGLLPADVP-VNESTG 279

Db 273 VEANREIELKQIK--IGQPVVYVD-AVNVEIKGVNDSFSFSTG--AFSPISATNATG 327

QY 280 QIT-----LRAAVPNDQ---NIMPGLYVAVLMD 305

Db 328 NPTKIVQRLPLRKILENQPDIKLRPGLSVAVVSD 363

```

RESULT 23
US-09-328-352-6273
; Sequence 6273, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6273
; LENGTH: 423
; TYPE: prt
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6273

```

Query Match	9.1%;	Score 184;	DB 4;	Length 423;
Best Local Similarity	23.5%;	Pred. NO. 1.3e-09;		
Matches 92;	Conservative 66;	Mismatches 178;	Indels 56;	Gaps 15

```

QY 12 AALAAV--AIVLSCGG-----GDAOGGQOPAGREAPVGVVTHPQVALL 60
Db 36 AAITAILGLLVNSKNKNSSEPEEGHAEHEEBEHHDSEKPLL--LTAQMOGONLK 94
QY 61 VE-----LPGRL--ESLTAIVRAQVCGIIQKRLQOESYVACQPLVYIDSS 106
Db 95 IEQAELEGVPOLOTYTPAKLVVNTDROAHVSPFSGRVEAVVYELQOVVKGOAL----AS 150
QY 107 TYEANLESARQOLATAQOTLAKADDLARYKPLVAAEVSROEYDAAVTAKESAEGVKA 166
Db 151 LLVPDLVDOQANLQIAQSNLELARDODYERESLMSQISAKODYORAVYAQOALQVKA 210
QY 167 AQAAIKSGI---NINRSRITAPISGFIQGSKVSCTLLMAGDTVLATIRQTNRYVN-- 222
Db 211 SRSRSISAGGSGSGRYTLTPPIAGIVSNKDIVGENVQADQ--LFTINLODQMLTF 268
QY 223 VTQSASEVMKLRQRIAEKGLLAADGVIAVGIKFDGQTVYPEKGRLLFPADVNNESGQIT 282
Db 269 ILPSNANINNVQNOIEFKSL-----QGNQIFSAQVQGLITE--ADAQGRQ 314
QY 283 LRAAVPNDQNTLMFGLVYRVLMDOYAVDNAFVPOQAVTR--GAKDTM--IVNAGGME 338
Db 315 VRAKVLNNSSELRLPMVNLVNLNGSQTQVTRVYKQAQVQVEGKQVIFTPKIVKTGEFE 374
QY 339 PREVTVAQO--QGTNMI--VTSGLKQGDKVVVG 368
Db 375 PVTVOLQGRSKDQGVVEVVKGINBSQRIYIAG 406

```

```

RESULT 24
US-09-252-991A-3115
; Sequence 3115, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18

```

PRIOR APPLICATION NUMBER: US 60/074, 788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094, 190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 31115
LENGTH: 517
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31115

Query Match 9.1%; Score 184; DB 4; Length 517;
Best Local Similarity 23.9%; Pred. No. 1.8e-09;
Matches 79; Conservative 52; Mismatches 162; Indels 38; Gaps 8;

51 TYNPCTVALTYLPG--RLESRTADVRAQVGGIIQKRLFGQSGVVRAGQPLVQIDSTY 108
195 TARTISLQVNSLPGEIRNEDRTAHVIRLPQIDVSPANIGAKQGEILLAVIS--- 251
109 EANLESABAQLATAQATLAKADADLARYKPLVAABAASVROEYDAVTAKRSAEACVKAQ 168
252 -PQLSDQSEFAAQRRLSLAQSTYKREQLMKEGISAQEFLLARQGLQAEIALLNAR 310
169 AAIKSAQINL-----NRSRTAPISGFGQSVSGTLLMAGDTTVALITRQTPMYNV 223
311 AKIALGQNPISLQGNRYELRAPFAGVLVEKHLTQCEPVDG--TANVFLLSLSVWATF 368
224 TQSAEWMKLRQIAEGKLLADGVIANGKPD-----DGVVPEKRLFPADVNMES 277
369 NVPA-----QLL---GGVRGSKVKVLAQALDSEV---EGVSYIGLLGEQ 409
278 TGOITLRAAVPNDQNLMLGVLVYRLMDQVAVNAFVVPQAVTRGAKDTVMIVNAQGM 337
410 TRAAARVTLSPSTWTRGLFVSQVVAETREKVLVTADGAVQVDGEDVVFVVAAGF 469
338 EPREVTVAQOQGTNMIVTSGLKDGKVVVEG 368
470 VVQPVKLGISDQGRVEVLEGLRAGSQVAASG 500

RESULT 25
US-09-252-991A-18501
Sequence 18501, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074, 788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094, 190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18501
LENGTH: 354
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18501

Query Match 9.1%; Score 183.5; DB 4; Length 354;
Best Local Similarity 24.6%; Pred. No. 1.1e-09;
Matches 86; Conservative 41; Mismatches 116; Indels 107; Gaps 12;

8 AMRAAALAAVALVSSCGKGGDAQGGPAGREAPAPVGVVTVHPOTVALTVELPGR 67
26 AVVVALGALAFWLS-----RDHQQDALRIYXGVDI--REVQALFQPGR- 70
68 ESLRTADVRAQVGGIIQKRLFGQSGVVRAGQPLVQIDSTYBANLESABAQLATAQATLA 127
71 -----VMQWAFDEGDVAVSAGARLALIDPQRYEALAAQAQVQVAAQAEIA 115

128 K-----ADADLARYKPLVAABAASVROEYDAVTAKRSAE 161
116 KLRGRLRQGETTQARREALRQALATETERNFQGSGLMSGATSQRVDAAARTADQA 175
162 AGVRAQAQAIAKSAGINLNRSHRTAPISGFGQSVSGTLLMAGDT-----TVLATI 213
176 AGVENAKAALISQASEGFEKEDIAAAEARLAAQAQAATTALADTELAPSSGTVIARV 235
214 RQTPMYNVNVTQSASEWKL-----RQIAEGKLLADGVIANG--IKFDGQ-TYYP 262
236 REPQSMV--VSQASVYSLDKVPYVRAYVESDL---GRIAGQTVVRVXSDSEKVV- 288
263 EKGRLFPADP-----VNESTQITRRAVP 288
289 -RGQIGFISPRAEFTPKVTETDRLTLVRLVRIVIDETSDRLRQPM 337

RESULT 26
US-09-328-352-7671
Sequence 7671, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7671
LENGTH: 356
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7671

Query Match 8.9%; Score 179; DB 4; Length 356;
Best Local Similarity 25.0%; Pred. No. 3e-09;
Matches 74; Conservative 45; Mismatches 83; Indels 94; Gaps 11;

70 LRTADVRAQVGGIIQKRLFGQSVYRAGPL-----YQIDSTYBANLES----- 114
60 IROYSLAFESQGRLOKLVQDEGDKVOGVQVATNTNALQIQAOQAQLAQOBAIVQ 119
115 -----ARQIATQATLAKADADLARYKPLVA-----EAVSROEYDAVTAKRSA 160
120 EVGARPEITTOAKAQLASQAQELDKTNKRLQRLQILVSTGRLAISQELDYAKSNQSHA 179
161 EAGVAAQA-----AIKSA-----GINLNRSHRTAPISGFGOS 194
180 DAAVERQANDELIIKARKEDREATRAQVEVTKANLDLIVNLTQALRSFVNA----- 234
195 KYSEGTLLMAGDTT-----VLATIRQTPMYNVNVTQSASEWKL-----ROI 237
235 -VVRARLOEVDMDTTQAQAVVTLALTNPKWIRVYVNEQDLSSIKWGSTAQIIRDAYPNOP 293
238 AEGKLLADGVIANGIKFPDGTVYPEKGR--LFPADVNESTQITRRAVPNDQ 291
294 MNKTI-----GIYISVAETPKVTQTEIRITLVY-----EVRVYVADPNDQ 335

RESULT 27
US-09-328-352-5910
Sequence 5910, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 5910
LENGTH: 358
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5910

Query Match 8.6%; Score 174.5; DB 4; Length 358;
Best Local Similarity 24.6%; Pred. No. 8.5e-09;
Matches 88; Conservative 50; Mismatches 139; Indels 81; Gaps 11;

OY 1 MAFVAFKMAAALAAVALVSSCGKGDAAGGQGP-----GREAPAPVGVVYH 53
DB 5 LCFPLILILIALYALVAFVFLNHSVSTDNAYVGAETQITQTSWSGVQAQVILKDTQTVH 64
OY 54 POTVALTEVLEPRLSTLTADVAQVGGIIQKRLFOE---GSYVAGQPLYOIDSSTYEA 110
DB 65 RGDV--LVIRIDRDAKIALAQAEALAK--AKQYKQTAANSSSLNSQVVRAD----- 114
OY 111 NLESAPQLATPQATLAQADDLARYKPLVAEAVSRQEDAAVTAKESAENG----- 163
DB 115 EINSKAOVQAQADYDRAALELNRAQLAASGAVSEKELTFQAQSAVETAKAGLELAKAG 174
OY 164 -----VKAAQAIAKSAGININRSRTAPISGFI 191
DB 175 LAQATSSRKAABSTLANEALIQVSESTPDVQAQHVQAQDLERTVIRAPVDGYI 234
OY 192 GQSKVSEGLTNAGDTVLATIRQTNPMYVNTQASAEWMLKLR--ROIAGKLLAADGYI 249
DB 235 TRRNIGQVQRAVPG--TSMNMIIVPLNDLYDANFESQKKVPRQPTLLSDLYGDVE 292
OY 250 AVG--IKFDGTVVPEKGRLLFADVNVSESTQITLRAAVPNDONIMPGYVRVLM 305
DB 293 YHGKVVSGSGGT-----GSAFALIPQONATGMWIKVQRLP-----VRIALD 334

RESULT 28

US-09-252-991A-25258
Sequence 25258, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR APPLICATION NUMBER: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 25258
LENGTH: 510
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25258

Query Match 8.6%; Score 174.5; DB 4; Length 510;
Best Local Similarity 22.7%; Pred. No. 1.5e-08;
Matches 99; Conservative 61; Mismatches 145; Indels 131; Gaps 19;

OY 8 AMRAAALAAVALVSSCGKGDAAGGQGPAGREAPVGVVYHPTQVALTEVLEPRL 67
DB 102 AQAVQALGLRTAEVH--GKIG-----ADVAVVGTIVLLNDRDSI----- 139
OY 68 ESLRTADVRAQVGGIIQKRLFOE--GSYVAGQPLYOIDSSTYEAESARAQATQAT 125
DB 140 -----VQARTAFVE--RVYARAAGDVIAAGAPL----- 166
OY 126 LAKADADLARKPLVA-AEAVS--RQEDAAVTA--KRSAGV--KAAQAIAKSAGINLN 179
DB 167 -----ADLLPKWIAQREFLSVAMGDAPLTAARQLLLILGMPOALLAQVETGEPKG 221

OY 180 RSRITAPISGFIQSKVSEGLTNAGDTV---LATI-----ROTNPMYVNTQSA 227
DB 222 LYVTTPQGLVIAELMVRQGTWVSAGESLTVNGLASVIAEAVPEAGSGPQVQV----- 276
OY 228 SEWKKLRQIAEGKLLAADGVIAVGIRKDDQTVPEKRLLFADPVNVSESTQITLRAAV 287
DB 277 -----QTRQVLAIFPGEVL-----QARIVSIIPEARDRTRTVKVRLEM 315
OY 288 PNDONIMPGYVRVLMDOVAVDN-----AFVPOQAVTR--GAKDTVMVINAQGWEPREV 342
DB 316 ANPQRLKAG-----MGQIALKNEQRPALLVPEEAVIRCKRLALAYVVDQGFHFHEV 370
OY 343 TVAQQGTMTWITVIGSLKDGKRVVEG-----ISAGITGAKKVTPEKMASSNQAAAP 395
DB 371 QVGAEEVGQVLVNGGLAAGQVVAQFLIDSEASLGVLP-----GAASAPASP 422
OY 396 QSGVQTASEACTASEA 411
DB 423 HNGHGTSSPTAAATNA 438

RESULT 29

US-09-252-991A-23319
Sequence 23319, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR APPLICATION NUMBER: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23319
LENGTH: 495
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23319

Query Match 8.6%; Score 173.5; DB 4; Length 495;
Best Local Similarity 20.5%; Pred. No. 1.8e-08;
Matches 97; Conservative 60; Mismatches 143; Indels 173; Gaps 15;

OY 23 SSCGKGDAAGGQGPAGRE-----APAPVGVVYHPTQVALTEVLEPRL---SLRTADV 75
DB 48 ATGARRTPAQQGPEDRSPRPAQPGRRPPTFPILRCARYRESAGRMNEAPMTTANS 107
OY 76 RAQVGGIIQKRL----- 88
DB 108 ETPAGNPKRRKWLILLAVILATLASVAMEFFYGRWHEDDTDAYINGNVQITPQIVGT 167
OY 89 -----QEGSVVAGQPLYOIDSSTVE-----ANLESARQAT 121
DB 168 VSIIGADGDLVRKQGLVRFPSDDADIALQREANLAHTVROVRLFSNVDDGRAEVAT 227
OY 122 AQATLAKADADLARYKPLVAEAVSRQEDAAVTAKESAENG----- 163
DB 228 RKVALAQAEDYKRRKLLADGALISQELAHARDALDSAKASLTSSEQOINTNRALVDT 287
OY 164 -----VKAAQAIAKSAGININRSRTAPISGFIQSKVSEGLTNAGDTTV----- 210
DB 288 QITSHDPVKAQAQRLQVAYDDARSTIVAPVTGVAKRSQVQVQVPGNAALAAVBLDQ 347
OY 211 -----ATIRQTNPMYVNTQASAEWMLKRLQIAEGKLLAADGVIAVGIFKDDGTVPEKGR 266
DB 348 ITWDANFKETQLKHMRIQGP-----VEIRSDLYSGDVARYSGTVSLGV-----GS 394
OY 267 LRFADPVNVSESTQITLRAAVPNDONIMPGYVRVLMDOVAVDNAFVVPQQAATRGAK- 325

Db 395 AFSLLPAQATNGNWKIVQRP-----VRHID-----PDELQHPRLI 433
QY 326 ----DTVMIVNAQCG-----MEPRE-----VTVAAQCGTMMIVTSGDKGDKV 365
Db 434 GLSMVVKVDLHDSGPAALAOQPPREALPSTVDYQOO-----LASADKLI 477
RESULT 30
US-09-328-352-7412
; Sequence 7412, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7412
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7412
Query Match 8.6%; Score 173; DB 4; Length 365;
Best Local Similarity 22.6%; Pred. No. 1.2e-08;
Matches 67; Conservative 55; Mismatches 104; Indels 70; Gaps 10;
QY 75 VRAQVGIIQKRLFOEGSVYRAGOLYQIDSTYEA-----NLE 113
Db 72 VMPKISGEVLEILLINDNQVKKGETLAVDHRDYQAQYDAHSVSLKEALGVQOOQEK 131
QY 114 SARAOATVATQATLAKADADLA-----RYKPLVAEAVSRQEDYDAAVTAKSABAGYKA 166
Db 132 SARSSITTEANGSVAAQADLAKLKEFERIYQDLKQGVITTRONFEGISQYITLQAQOLSK 191
QY 167 AQA-----IKSAGINLN-----RSRTAPISGFIGOSKVSE 198
Db 192 AQAANAALAEAGLSLQASRAQLADIQSANANLNYVDLASSKVSVPVSGKISGLATQK 251
QY 199 GTLLNAGDTTLATIRQINPMYVNTQASAEVWKLRRQIAEKKLLAADGVIAVGIFPDG 258
Db 252 GSRVSP-QTRLMALIIPE-NSLYQANFKETQIEKM--HIGQVKKLAYAPSLNFTGKIE 307
QY 259 TYPEKGRLLFPADVNESTG-----QITLRAVPNDQNI--LMPGLYVRLMD 305
Db 308 SFPSPASGATFSLMPDN-ATGNFNKVVORIPVRIADSSPHIDLKPKMSVATVD 362
RESULT 31
US-09-328-352-4722
; Sequence 4722, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4722
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4722
Query Match 8.5%; Score 171; DB 4; Length 386;
Best Local Similarity 22.9%; Pred. No. 2.1e-08;
Matches 83; Conservative 56; Mismatches 120; Indels 104; Gaps 13;
QY 37 PAGREAPAPVGVVTHPQTV--ALTVELGRLESL-----RT 72

Db 31 PAPAPAPASKLLPTKRSSTLLMGLVILITVLAAMRIGPATSVQOQDINSYVKGKT 90
QY 73 ADVRAQVGIIQKRLFOEGSVYRAGOLYQIDSTYEA-----ANLESARAOATATAQAT 125
Db 91 TILSSQINGVKVDLVKDFHVKKGQVLMHIDATTTIDQKVAQAASGVQAOKTILANQOS 150
QY 126 LAKADAD-----LARYKPLVAEAVSRQEDYDA-----VTAK 157
Db 151 IAKQADIVAAQAVEQVRAQYELSLAQLRHYQOLGNSGASKEODPAAADENNLAL 210
QY 158 RSEANGVAAQAALIKSAGI-----NLNSRTAPISGFIGOSKV 196
Db 211 KQAEANVIVAEALKTAQVABAGEAQSAAQLDQATTQKDYSVIAPMDGQGEVNP 270
QY 197 SEGTLNAG-----DTVLATIRQINPMYVNTQSA--SEWKLRRQIAEGKLLA 244
Db 271 RVGQVVAAGSQLILIPQOTVIANFKETQIANRRIGQKAMFTYDAMKHKFTGHVEQIS 330
QY 245 ADGVIAVGIKFDDGTVYPEKGRLLFPADVNESTGQITLRAVPNDQNI--LMPGLYVRV 302
Db 331 P-----AAGSEF--SVLKPDNATGNFT--KVQGRIVRITID--PNOEGIEHLRPMQSVIT 380
QY 303 LMD 305
Db 381 SVD 383
RESULT 32
US-09-328-352-8202
; Sequence 8202, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8202
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8202
Query Match 8.3%; Score 168; DB 4; Length 336;
Best Local Similarity 22.3%; Pred. No. 3.3e-08;
Matches 69; Conservative 54; Mismatches 106; Indels 80; Gaps 10;
QY 75 VRAQVGIIQKRLFOEGSVYRAGOLYQIDSTYEAANLESARAOATATAQATLAKADADLA 134
Db 54 VSSDVAGLVTEVLVODNOTVKKGOVLFPKIDVSRALDVEQAKSLAKKAAYQAQAGLA 113
QY 135 -----RYKPLVAEAVSRQEDYDAVTA-----RS 159
Db 114 QAKANLKSSTNIKLAENKARYSNLMDG-AISKQEQOVFATRDQSHAEHQLQAALQ 172
QY 160 AEAGVKAQAALIKS-----AGINLNSRTAPISGFIGOSKVSEGLTLNAGDTT 208
Db 173 AEATIKQOQALVEATSNLHLAEINMHRAAVAAVADGTLNFPDRLRPGYVVGQAVALL 232
QY 209 -----VLATIRQINPMYVNTQASAEVWKLRRQIAEKKLLAADGVIAVGIFPDGYVP 262
Db 233 DRKQLYVGVFEETKLRIRIHIGDEATYQOLMDRQIKIGHVG-----IASGIEDERS--- 285
QY 263 EKGRLLFPADVNESTGQITLRAAVPNDQNIIMPGLYVRVIMDQVAVDNAAFVPPQAVTR 322
Db 286 SSSKLL--ANVNPTFSWVRLAQRVP-----VKIYLDAPKX-----ELAFVS 325
QY 323 GAKDTVMIV 331
Db 326 GRTATVTHII 334

RESULT 33
US-09-252-991A-20336
; Sequence 20336, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20336
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20336

Query Match 8.1%; Score 164.5; DB 4; Length 314;
Best Local Similarity 25.4%; Pred. No. 6.6e-08;
Matches 75; Conservative 41; Mismatches 118; Indels 61; Gaps 10;

QY 74 DVRAVGGIIOKRLFOEGSYVRAGQPLVQIDSTYEANLESARQATATKADADL 133
DB 59 NVADVSGIVDVPRDQVLKQGLDLPDHRIVAKAESLVSRKATLMMROLN- 117
QY 134 ARYKPLVAEAVSRQYDAATVAKSAEAGVKAQAQAISAGINLNSRITAPISGFIQ 193
DB 118 ARRAEDENWVSHESRDAHTAAAMADYEOKAQDADARLDERVVAQVDGYVTN 177
QY 194 SKVSEGTLLNAGDT-----TVLATRQTNPMYVNTQSASEVMKLR 235
DB 178 LNVHGDVARVGEAKMAVIDKNSYVWVGYFETKLPYRREGDPVDM----- 223
QY 236 QIAGKLLAAD-GYIVAVIKRDDGTVPEKRLFPADPVNVESGQITLRAVPNDONIL 294
DB 224 QLMGSEHLKGVESIARCIYDRD---PE-SRELTAD--VNPITNWRLAQRVP----- 271
QY 295 MPGLYVRVLMQVAVDNAFVVPQAVTRGAKDTVIYVNAQSGMREBVTVAQOQG 349
DB 272 -----VRVHIDE-----VPDGVLSAGITCTIVAKPGG--RDDQSAHAQAPG 311

RESULT 34
US-09-252-991A-17970
; Sequence 17970, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17970
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17970

Query Match 8.0%; Score 161.5; DB 4; Length 357;
Best Local Similarity 26.0%; Pred. No. 1.6e-07;

Matches 69; Conservative 40; Mismatches 101; Indels 55; Gaps 10;
QY 71 RTADVRA-----QVGGIIOKRLFOEGSYVRAGQPLVQIDSTYEANLESARQATATQ 123
DB 107 RDAVRADVIVVADVSGVWTDLEVKDQNVKQVDVLMRIDQERYQANLEQARRVAETRH 166
QY 124 ATLAKADADLARYPL---VAEAVSRQYDAATVAKSAEAGVKAQAQAISAGINLN 179
DB 167 QOYLLRQNEAARRSRILGASISADKENAQINAI-----ARSEYQEAQVKTAEINLK 221
QY 180 RSRTAPISGFIQSKVSEGTLLNAGDTYVATIRQTN---PMYVNTQSASEVMKLRQ 236
DB 222 RSELRAARNGQVNLRLAQGNVATAGQ-AVVALVDQGSFYVVAAYFEETKLPGIRGWRQ 280
QY 237 IABGKLLAADGVIVAGIKFDGTVYPEKRLFPADPVNVESGQITLRAVPNDQ---NI 293
DB 281 V---RLMSGDPT-----DGT-----VESISGTTDRNSTPQQQLANV 316
QY 294 -----LMPGLYVRVLMQVAVD 310
DB 317 EPTFNWVRLAQRIPVRIRLDQVPAD 341

RESULT 35
US-09-252-991A-23589
; Sequence 23589, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23589
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23589

Query Match 8.0%; Score 161.5; DB 4; Length 399;
Best Local Similarity 24.1%; Pred. No. 1.9e-07;
Matches 92; Conservative 54; Mismatches 144; Indels 91; Gaps 15;

QY 11 AALAAVAVALVSSCGKGDPAAG-----GQPA-GREAPAPVVG-----VV 50
DB 23 AAQOPAAALRRVAAGLARSAPVPRVVLRLDRSDPVDGVDVPMNLQGMPLRLATLLAV 82
QY 51 TVHPQVATLVELPGRLESRTADV-----AOVGGIIOKRLFOEGSYVRAGQPLVQI 103
DB 83 LVAAALVFLAMEYTRTPWRDARVADVTLISADVGLITRLPLSNQVYKELLSTI 142
QY 104 DSSTYE-----ANESARAQATATKADADLARKPLVAEAVSRQ 149
DB 143 DPARYELAVLARAVVAVARPAALGESRPALEASQAQIRQRSEQRORSALQKSMLSVEB 202
QY 150 YDAAVT-----AKRSAEAGVKAQAQAISAGINLNSRITAPISGFIQSK 195
DB 203 NEKAQDVTSLAQAEILRNQASLGLAQANVELAEHALQQAQDLERTVRAVPSCGYVNLQ 262
QY 196 VSEGTLLNAGDTYVATIRQTNPMYVNTQSASEVMKLR-RQIAEGKLLAADGVIVAGIK 254
DB 263 TREGDYAHAG--VPLALVDRDSFYVSGYFETKLPQIRVGSRAVELMS-----GER 313
QY 255 FDDGTYP-----EKGRLLFPADPVNVESGQITLRAVP-----NDONITM 295
DB 314 F-DGRVQSIHAATITRENAESRLAN--INPSYTWKLAQRIPVRIALDPAYRORNTLR 370

```

QY      296 PGL--YVRVL--MDQVAVDNA 312
          | : ||| | ||:|
Db      371 AGVTATIRVLPAEDTAKVDSA 391

```

```

RESULT 36
US-09-252-991A-28902
: Sequence 28902, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 28902
: LENGTH: 373
: TYPE: PRN
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28902

```

Query Match	8.0%;	Score 161;	DB 4;	Length 373;
Best Local Similarity	24.0%;	Pred. No. 1.9e-07;		
Matches 61;	Conservative 44;	Mismatches 77;	Indels 72;	Gaps 7

```

QY 75 PRAOVGGIIIOKLFQEGSVVRAGOPLOYIDSSITVEANLSARQALPDQAOTLAKADADA 134
Db 78 VASQIGARVEEVLVRDNGHVDKQOLLVRLLEDADFKLAVERAQAALTBRELAQAASKLV 137
QY 135 RYKPLVAEA-----VSROEYD----- 151
Db 138 OOGSLIAASADVNASQATLIGRAQIDILNREALRKPGVSEERKVTLLTDNHHVARSQALAK 197
QY 152 ---AAVTAAR-----SAAEGVKAQAALIKSAGINUNRGRITAPISGFQGSQKVE 198
Db 198 ARAIDLEAQRVQNDTLGAETIKRLIEAQJASARTELQAQIEINLSRLLHSPISGLVGQSSARN 257
QY 199 GTLLNAGDITVLATIRQTNPMYVYNTQASAEVWKLRQRQAIEAGKLLAADGVIAVGIKFDDG 258
Db 258 GQYVGVG--THLLSLVPEDEDIWWQANFKETQYGBMR-----DGQKA-RLTEDAF 303
QY 259 TYYPEKGRU--LFA 270
Db 304 PDTPIDGRIDS.LFA 317

```

```

RESULT 37
US-09-328-352-5275
; Sequence 5275, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5275
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5275

```

Query Match	7.8%;	Score 157.5;	DB 4;	Length 374;
Best Local Similarity	19.6%;	Pred. No. 4.2e-07;		
Matches 72;	Conservative 70;	Mismatches 172;	Indels 53;	Gaps 7

```

Oy 1 AVAVLVISOGCKGGDAAGOGQPRAGRAPRVWG-----VVTHPQTVATVYELPG 65
Db 20 SUBSILIACNO-----QVEVPTTASPEKKIEILIPQDILPKESGSLAQTVFTG 67
Oy 66 RLESRTADVRAOVGGIIQKRLFQESYVRAGQPLYOIDSSTVEANLESARQATQAAT 125
Db 68 TIRAVOGSSIQAOVSATATAVTADVGQKQGVULVRINQDNMAARLAQAOANLASAOQ 127
Oy 126 LAAQADDLARYRPLVAEAENVSRQOEIDAAVTKRSEAGVYKAAQAAISAGINLRSRITA 165
Db 128 AELARNLMNRKQRLINQGRARAEVEEQSVYKQQLBESVRAQQAANVBIAKKABDGIITS 187
Oy 186 PISGFIGOSKVSBEGLTLAGDFTTVLATIRQTNPMVWVNTQASAEVMLKRLROIAEKLAA 245
Db 188 PISGVITTKQVEBGQTVSVGT---LFEIYNPDQLEIQAR---LPIEQOSA----- 232
Oy 246 DGVIYAVGIRFEDDGYVPEKGR-----LFLADPVYNESTQITLRAAVPNDONIMPLG 298
Db 233 -----LKVGSISIQYQIQGNSKQHLALITRISPVADQDSRQIBFPASPKXAISLSIGA 285
Oy 299 YRVLVMDQVAVDNAFVVPQAVTRBAKQT-VMIYVAAQGMREPREVYVAAQOQGTWIVTSG 357
Db 286 FINGIILHNDNSOGQTIPLDSIQNLQHPDFFWVIRINQTIQKXKRVREORVENEMIALVOG 345
Oy 358 LKXGDKV 364
Db 346 LQSTTDV 352

```

RESULT 38
 US-09-252-991A-28283
 ; Sequence 28283 Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 28283
 ; LENGTH: 449
 ; TYPE: PR1
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-28283

Query Match	7.7%	Score 155.5;	DB 4;	Length 449;
Best Local Similarity	22.0%;	Pred. No. 8.9e-07;		
Matches 74;	Conservative 54;	Mismatches 113;	Indels 95;	Gaps 11

```

OY 6 LPRLBSLR-----TADVRAQVGGIIKRLRFQEGSVYRAQCPYXIIDSSRY-----108
Db 126 LPRLPRLPPQQLTRPVYVYARVVSQVAVLVSNNGHVQPEVLFRLDPEEFLAVYAOE 185
OY 109 -----EANTLESAPQQLATAQATLAKADADARARYKPLVAAEAVSROEYDAVT 155
Db 186 LALPEARNTNRELDAALASAKADLLAARSAGELDSEARPTAQVVRHHVSOQHGVSA 245
OY 156 AKRSAAAGYKAAQ-----AASAGINLRSGITPISGFI 191
Db 246 QQAQAARAPYAAQAARIGELTARGTAGEDNLRQARNALAAQALQLOYSVSRVDRRGTL 305
OY 192 GQSKVSEGTLLNAGDT-----VATIRQTPMVPVANTOSASEMVKLRQIARGL 242
Db 306 SNQLRLRGTYVAGTGVVALVDNRIDYIADPREKSLRYVRGPDAAVVFQARPEVEFGAR 365
OY 243 LAADGVIAVGIKFDGCTVYPEKGRLLPADPVVNESGTOITLRAAVPNDONILMGLYVRV 302

```


THIS PAGE BLANK (USPTO)

PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL05732.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 11679; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL16511), expressed DNA
CC sequences (ABL16140-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1201 AA;
Query Match 6.6%; Score 133; DB 22; Length 1201;
Best Local Similarity 20.6%; Pred. No. 0.019;
Matches 96; Conservative 41; Mismatches 210; Indels 118; Gaps 15;
OY 9 MRRAALAAVALVSSCGKGD--AAOGGAPGRAPVGVVTHQVTAALVELPGR 66
DB 786 LATGATSAAPLASSINATATATVAAATKSPPAAGVAPTLVGIAAT-CTTAAANAGKAT 844
OY 67 LE-SLRTDVRARQVGIIOKRLFGESYVRAGQPLQYIDSSYEAHLSARALTAQAT 125
DB 845 LKTTISSGAVATAAG-----GNKAASVAAATAT 873
OY 126 LAKADADLARYKPLVAABAVSRQEDAAVTAKSAAGVKAQAQAIKAGINLNSRTTA 185
DB 874 VAAKTPPAAAMLVAA-SVKVSGTDKAPLKTTSASVAVPBGTKAS-----TA 923
OY 186 PISGTTGGSKVSEGLLNAGDTTVALTRQTP-----MYVNTGASSEVMKLRQIAGCK 241
DB 924 PSVATAVAVSVSSGATVPAAASKAAPLKATAPPSLAAVNATQAAKST-----FVAAS 978
OY 242 LLAAG-----VIAVGKFPDGTVPPEKRL 268
DB 979 ALVAPAPPAKSTASVAKTNTNAGSKPTATLSTCKSDAPAAASGVNTNGKTPSGGATS 1038
OY 269 FADPVNVESTGQITLRAAVPNDONILMPGLYVRVLMDOVAVNATVFPQAVTRG----- 323
DB 1039 TAQPAVNG---KATAAAAAQSQ-----GATATATPTTALNOKTIVSAGSGKDVPS 1090
OY 324 -----AKDTVMIVNAOGMEPREVTVAOOGTNNIVTSGLKDGKRVVE--GI 369
DB 1091 RSTDPAPPAKGAASNISAKSTKAGTKAQTONETSPTKSAVK--PIKAGSTSPTRAVVSKGCA 1149
OY 370 SIAGITGAKKVTPEKMA---SEENQAAAPOSQVOTASEAKTASEA 411
DB 1150 SAAKTAASAAATATATTAAGSSSSNKKAATSTATATGSOAKTKGNA 1194
RESULT 40
ID AAY38881
XX AAY38881 standard; Protein; 392 AA.
AC AAY38881;
XX
XX 08-OCT-1999 (first entry)

XX
DE Neisseria meningitidis antigen encoded by a partial ORF5.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN W0924578-A2.
XX
XX 20-MAY-1999.
PD
XX
XX 09-OCT-1998; 98WO-1B01665.
PF
XX
XX 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024180.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
PI
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
DR N-PSDB; AA212303.
XX
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 427; 524pp; English.
SQ Sequence .392 AA;
Query Match 6.6%; Score 132.5; DB 20; Length 392;
Best Local Similarity 31.0%; Pred. No. 0.0042;
Matches 39; Conservative 17; Mismatches 55; Indels 15; Gaps 3;
OY 262 PEKGRLLFADPV-----NESTGQIT-----LRAAVPNDONILMPGLYVRVLMDOV 307
DB 251 PIKAKLSDVDGLTTMSSGGNSSTDTASNAVYYAASFVNPOGKLATGTTONTYVID 310
OY 308 AYDNAFVVPQAV-TRGAKDTVMIVNAOGMEPREVTVAOOGTNNIVTSGLKDGKRVV 366
DB 311 GVKVLLIIPSLTVANRGKAFVRVLGADGKAEREIRTGMDSMNTEVKSGLKSGDKVI 370
OY 367 EGISIA 372
DB 371 SEITAA 376
Search completed: September 8, 2003, 14:00:46
Job time : 47 secs

DR N-PSDB; AAH81443.
XX
XX Novel nucleic acids encoding proteins required for *Escherichia coli*
PT proliferation, useful for screening for antimicrobial agents -
XX
XX Claim 19; Page 554-555; 596pp; English.
XX
XX The present invention describes a purified or isolated nucleic acid
CC sequence (1) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (1) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (1) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.
CC For example, nucleic acid probes complementary to proliferation-required
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the *Escherichia coli*
CC proteins given in AAH89239 to AAH89431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.
XX
XX Sequence 355 AA;
SQ
Query Match 6.9%; Score 138.5; DB 22; Length 355;
Best Local Similarity 26.3%; Pred. No. 0.0011;
Matches 52; Conservative 32; Mismatches 55; Indels 59; Gaps 6;
OY 65 GRLESIRADYVRAQVGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAOLQA 124
DB 41 GRIETE-VDIASKAGRIDTIVKEGKRVREGLAKMDTRVLOEQLALIAQIKKQOS 99
OY 125 TLAKADADL-----ARXKPLVAEAVSROEYD----- 151
DB 100 AVAAQAQLLEOROSFTRAQSLVNRQAELEDSVAKHRRSLAQRAISAQQLDDRAA 159
OY 152 --AANTARSAAGCYKAAQAIAKSGIN-----LNRSRTAPIS 188
DB 160 AESAQALESAAQVSAKAAIEARTNIIQAQTRVEAQAOTERRIADIDDEIKABRD 219
OY 189 GFIGQSVSE-GTLNAG 205
DB 220 GRV-QRYVAEPGEVLAAG 236
RESULT 38
AAVS5920
ID AAVS5920 standard; Protein; 444 AA.
XX
XX AAVS5920;
AC
XX
XX 15-FEB-2000 (first entry)
DT
XX
XX Pseudomonas fluorescens ABC transporter cassette component B.
DE
XX
XX ATP-binding cassette; transporter; operon; lipBCD; *Serratia marcescens*;
KW microbe; protein secretion.
XX
XX Pseudomonas fluorescens.
OS
XX
XX JP11276172-A.
PN
XX
XX 12-OCT-1999.
PD
XX
XX 27-MAR-1998; 98JP-0080597.
PF
XX
XX 27-MAR-1998; 98JP-0080597.
PR

XX
XX (TANA) TANABE SEIYAKU CO.;
PA
XX
XX WPI; 1999-626936/54.
DR
XX
XX N-PSDB; AAZ22700.
PT
XX
XX An ABC transporter gene - used to increase the ability of a microbe to
PT secrete proteins
PS
XX
XX Claim 2; Page 16-17; 28pp; Japanese.
XX
XX This sequence represents an ATP-binding cassette (ABC) transporter
CC component B from an *Pseudomonas fluorescens* strain 33 and encoded by
CC an ABC transporter operon. The operon comprises 3 genes where the
CC termination codon of the first gene overlaps with the initiation codon
CC of the second gene. The sequences of the encoded proteins have 60, 44
CC and 46% homology respectively to the lipBCD proteins from *Serratia*
CC *marcescens*. The novel gene and protein can give or increase the
CC ability of a microbe for secreting a protein.
XX
XX Sequence 444 AA;
SQ
Query Match 6.8%; Score 137.5; DB 20; Length 444;
Best Local Similarity 22.8%; Pred. No. 0.0018;
Matches 79; Conservative 65; Mismatches 103; Indels 99; Gaps 17;
OY 60 TVELGRLESIRADYVRAQVGIIQKRLFOEGSYVRAGOPLYQIDSTYEANLESARAOL 119
DB 58 TVVSGKRKAVQTLS-----PGVSRILVREGEAVKOGQPLFRDQONQADVHSLQAOY 112
OY 120 ATQAOTLAKADADLARYKPLVAEAVS-----ROEYDAVTAKRSAGAVK 165
DB 113 RMAVASVARQMSERDNOSTITFPALSLGNPDALALVLEGRQLFSSRBAFAEQAQIR 172
OY 166 AQAQATISAGININ-----SRITA-----PI--SGFTQ-----SK 195
DB 173 --ANIGATRAQLQGMRRASDLTAQASLRDQNLNLPADNGYIPRNLMEXYROLQSO 229
OY 196 VSEGTLLNAGDT-TVLATIRQTNPMYVNVTSASEVAK-LRQIAEGKLA----- 244
DB 230 VQDQLAQNTGSGVBEQALLESR---LKDQHSREYQKVRSLADQALSITIEQQLTS 286
OY 245 -----ADGVIAVGKIFDGTIVYPEKGRLLFADPVNVSTGQITLRAAVENDQ 291
DB 287 AGFDLQSEINAPADG-IAVNL-----SVHTE-GAVVRA-----GE-TLLEIVPQDT 330
OY 292 NILMPGLYKRVLMQVA-----VDNAFVVPQQAATVRGAKDTVMIVNA 333
DB 331 RLEVEGRPLVPHLVKGTHTLPVDTLFTAFNQSRTPRVPGEVSLISA 376
RESULT 39
ABB61629
ID ABB61629 standard; Protein; 1201 AA.
XX
XX ABB61629;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX *Drosophila melanogaster* polypeptide SEQ ID NO 11679.
DE
XX
XX *Drosophila*; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX *Drosophila melanogaster*.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 645 AA;

Query Match 7.5%; Score 151; DB 22; Length 645;

Best Local Similarity 26.5%; Pred. No. 0.0002;

Matches 63; Conservative 43; Mismatches 96; Indels 36; Gaps 8;

OY 27 KGGDAAAGGQAPARAPAPVVG-----VTVHPQTALVLPGLRLSLRTADVRA 77

Db 361 RGPKHGQBPQ--NRAPPELDAGELSWRTDHIISIAFGKMFSTGENMGRFFPFNTA--W 414

OY 78 QVGGIIQRLPOGGSYV---RAGGPVYQIDST-----YANLESARAQATQA 124

Db 415 QWGEREEQAINNEGDMITSPGRDTSNADKLNNNTAGAEENOIKEVATLMELRAQQA 474

OY 125 TLAKADADLARKPLVAABAVSRQEDAAVT--AKRSAAG-----VKAQAARISAGIN 177

Db 475 ELKLAARVYSKQQLAKQKAVSQQLDTRATPMAYKQKQIGTIDQIKRNOASLDTAKTN 534

OY 178 LNRSRITAPISGFIQSKSVSG-TLLNAGDTTVALTIRQTNPMVYVNTQSASEVWKL 234

Db 535 LDYTRIVAPMAGEVYQITTLQGGTIVTAAQQAENITLADMMSMLVKQAVSEADVHTK 592

RESULT 36

ABG17838 ID ABG17838 standard; Protein; 684 AA.

XX AC ABG17838;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #17829.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HSE-) HSEBQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PDB; AAS82025.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity -

XX PS Claim 20; SEQ ID No 48197; 103pp; English.

XX XX The invention relates to isolated polynucleotide (I) and

XX CC polypeptide (II) sequences. (I) is useful as hybridization probes,

XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX CC and gene mapping, and in recombinant production of (II). The

XX CC polynucleotides are also used in diagnostics as expressed sequence tags

XX CC for identifying expressed genes. (I) is useful in gene therapy techniques

XX CC to restore normal activity of (II) or to treat disease states involving

XX CC (II). (II) is useful for generating antibodies against it, detecting or

XX CC quantitating a polypeptide in tissue, as molecular weight markers and as

XX CC a food supplement. (II) and its binding partners are useful in medical

XX CC imaging of sites expressing (II). (I) and (II) are useful for treating

XX CC disorders involving aberrant protein expression or biological activity.

XX CC The polypeptide and polynucleotide sequences have applications in

XX CC diagnostics, forensics, gene mapping, identification of mutations

XX CC responsible for genetic disorders or other traits to assess biodiversity

XX CC and to produce other types of data and products dependent on DNA and

XX CC amino acid sequences. ABG00010-ABG30377 represent novel human

XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 684 AA;

Query Match 6.9%; Score 140; DB 22; Length 684;

Best Local Similarity 53.6%; Pred. No. 0.002;

Matches 30; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

OY 90 EGSIVRAGPPIYQIDSTSEYENLESARAQATQAOTLAKADADLARKPLVAABAV 145

Db 102 EGSVDVQAGSLYQIDPATYQANDYSAKGELEKASAAAIAHLTVKRYVPLVAKNEV 157

RESULT 37

AAG98387 ID AAG98387 standard; Protein; 355 AA.

XX AC AAG98387;

XX DT 21-SEP-2001 (first entry)

XX DE Escherichia coli protein sequence SEQ ID NO:435.

XX KW Escherichia coli; identification; proliferation; microorganism;

XX antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;

XX bacterial growth inhibition.

XX OS Escherichia coli.

XX PN WO200148209-A2.

XX PD 05-JUL-2001.

XX PF 19-DEC-2000; 2000WO-US34419.

XX PR 23-DEC-1999; 99US-0173005.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Forsyth RA, Ohlsen KL, Zyskind JW;

XX DR WPI; 2001-457376/49.

Db 346 DEQKAKKEVSLGNADAENQETSGLTNGAKVI 379

RESULT 34
ABU01173
ID ABU01173 standard; Protein; 399 AA.
vv

AC ABU01173;

DT 11-FEB-2003 (first entry)

S. pneumoniae type 4 strain protein from coding region #746.

KW Bacterial meningitis; pneumonia; sepsis; otitis media;

XX
NM
RM
XX

adulatory; gene therapy; vaccine.

Streptococcus pneumoniae type 4 strain.

XX 13
3W-170 / 30030M

XX

XX

XX

PA (GENO-) INST GENOMIC RES.

PI Massignani V, Tettelin H, Fraser C;

DR WPI; 2003-040579/03.

XX

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or ear infection -

PS Claim 1; SEQ ID No 1492; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pubd/published_pct_sequences.

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
CC

XX Sequence . 399 AA;
SQ

Query Match	7.5%;	Score 151.5;	DB 24;	Length 399;
Best Local Similarity	24.9%;	Pred. No. 9e-05;		
Matches 98;	Conservative 64;	Mismatches 163;	Indels 69;	Gaps 18

QY	12	AAALAAVALVSSCKGDA--AQQGAPAGAEAPAPVGVVYTHPQVATVLTPEGLLES	69
Db	15	AAAGAASVVL---GAGGILLFROPSPQALDEP---THLVAKEGSVASSVLLSGTVTA	68
QY	70	LRTADVRAQVG-GIIQKLFQEGSVYRAGQEPLOYIDSTYFANLESRAQLATA-----	122
Db	69	KNEQYVYFPASGDDDELIVSVGDKVSSGALVYKSSSEAOAAVDSRAVARADRHINE	128
QY	123	--QATLAKADADLIARY-KPLVAAEVSHOEYDAAVTAKRSAGCVKAAQAIIKSAGINLN	179
Db	129	LNRARNEAASAPAPQLPAPVGEGEDATVQSPPIVAGNSVASIDAQIGARPARADAAQLS	188
QY	180	RSRLTPISGIFGQSKVSEGLTLNAGDTTVLATIRQTNPMVNTQS---ASEPM-----	231
Db	189	KA-----QSOLDATTVL---STLEGTEVENVS--NVSKSPGASQVMVHIVS	230
QY	232	---KLRLQIAEGKLIADGVLAIVGIRKEDDETVPPEK---GRLLF-----ADPV	273
Db	231	NENLTQYKGSLSLYNL--ANLSVQGEVSR-TSKYVLPDKMTGSKLSYIDYPRKNNGEAASPA	287
QY	274	VNESTG-QITLRAAVPNDONILMPGLVYRVLMDOVAVDNAFVVP-QQAVTRGAKDTVMIV	331
Db	288	AGNNTSKSPYITIDVTGVEGDLKQGFVNI--EYKSKTKAILIVSVSLVMDSDKNYMWIV	345
QY	332	MAQGMREPREVTVAQQQGTNMIIVTSGLKDEGRKV	365
Db	346	DEQQRKKKVEVSLGNADANEQELITSGLNGAKVI	379

RESULT 35
ABG21616
ID ABG21616 standard; Protein; 645 AA

AC ABG21616;

DT 18-FEB-2002 (first entry)
xx

De Novel human diagnostic protein #21,607.
XY

human; chromosome mapping; gene therapy; forensic
 food supplement; medical imaging; diagnostic genetic disorder

OS Homo sapiens.

PN WO200175067-A

PD 11-OCT-2001.

PF 30-MAR-2001;

PR 31-MAR-2000; 2000US-0540217.

XX

2 XX

XX 0001 00000000/00000000

DR N-PSDB; AAS85803.

PT New Isolated polynucleotide

PT responsible for g

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
XX Claim 20; SEQ ID No 51975; 103pp; English.
XS

PS Claim 20; SEQ ID No 51975; 103pp; English

XX	18-MAR-2002; 2002WO-GB01305.
PF	
XX	22-MAR-2001; 2001GB-0007234.
PR	23-MAR-2001; 2001GB-0007360.
RR	
XX	(ISIS-) ISIS INNOVATION LTD.
PA	
XX	
PI	Herbert MA, Deadman ME, Hood DW, Moxon ER;
XX	
DR	WPI; 2003-029913/02.
DR	N-PSDB; AAD47834.
XX	
PT	New virulence gene from Hemophilus influenzae, useful for producing
XX	vaccines or antibiotics for preventing or treating pneumonia -
PS	Claim 4; Page 65-66; 132pp; English.
XX	
CC	The present invention relates to Hemophilus influenzae virulence genes
CC	and proteins encoded by them. The microorganisms or the peptides of
CC	the invention are useful for manufacturing a medicament for treating
CC	(veterinary) or preventing a condition associated with H. influenzae
CC	infection, particularly pneumonia or for identifying an antimicrobial
CC	drug. Sequences of the invention are useful in the production of vaccines
CC	or antibiotics to prevent or treat H. influenzae infection. They are
CC	also used in gene therapy. The present sequence is H. influenzae
CC	emwA protein (multidrug resistance protein A).
XX	
50	Sequence 390 AA.

Query Match	7.8;	Score 156.5;	DB 24;	Length 390;
Best Local Similarity	21.6%;	Pred. No. 3.2e-05;		
Matches	71;	Conservative 115;	Indels 91;	Gaps 10;

RESULT 33
AAV81743
ID AAV81743 standard; Protein; 399 AA.
XX
AC AAV81743;
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID49.
XX
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicemia; sinusitis; meningitis; therapy.

XX Streptococcus pneumoniae.
OS
XX
XX WO200006738-A2.
PN
XX
XX 10-FEB-2000.
PD
XX
XX 27-JUL-1999; 99WO-GB02452.
PF
XX
XX 27-JUL-1998; 98GB-0016336.
PR
XX 19-MAR-1999; 99US-0125329.
XX
XX
PA (MICR-) MICROBIAL TECHINICS LTD.
XX
XX
XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
PI
XX
XX WPI; 2000-195301/17.
DR
XX N-PSDB; AA291839.
XX
XX Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections
XX
XX
PS Claim 2, Page 53; 76pp; English.

Query Match	7.5%	Score 151.5	DB 21	Length 399
Best Local Similarity	24.9%	Pred. No. 9e-05		
Matches 98	Conservative 64	Mismatches 163	Indels 69	Gaps 18

QY 332 NAQGMEPREVTVAQQGGTNWIVTSGLKDDGKVV 365
: | : ||: | : ||| : | ||:

CC aluminum-resistant microorganisms. The products of the invention can be
 CC used to impart aluminum resistance to other organisms, particularly
 CC plants or microbes, especially to allow them to be grown in presence of
 CC aluminum at normally toxic levels, e.g. in acidic soils. This sequence
 CC represents the Acidiphilium cryptum Alub proteins described in
 CC specification.

XX Sequence 262 AA;

Query Match 7.8%; Score 158; DB 23; Length 262;
 Best Local Similarity 26.7%; Pred. No. 1.3e-05;
 Matches 67; Conservative 39; Mismatches 113; Indels 32; Gaps 8;

QY 7 KAMRAALAAVALVLSGCGKGDAAQGGPAGREAPVVGVT-VHPQTVALTVELPG 65
 DB 29 RRVRGAVLAALAV-----GVAALAIAGMTVG--APPVHVHTAKVGGDITRAVATG 81
 QY 66 RLESRTADVRAQVGGITQKLFQSGSYVRAGOPLYQIDSSYEANLESARQALTAQT 125
 DB 82 TVNPILTVYVGSVAGIITQDISCDVNTITVKLGQICARIDPRPYQAVVDQRAAVAVDQAQ 141
 QY 126 LAKADADI-----ARYKPLVAEAVSRQEDVAVTAKRSNAG-----VKAQAQAI 171
 DB 142 LRKDEASLDYARGNTRRLASLVPORYASENADSAKATTRQLEAOTLADRAVITRODEAKT 201
 QY 172 KSAGINLRSRITAPISGFIGOSKVSSEGLTNAGDPTVLTIRQTNPMVYVNTOSASEVM 231
 DB 202 AAAQNLGVTDTIVSVNGIVSERNVTIG-----QTVAAAP-QPTPLFLATDITR--M 251
 QY 232 KIRROIAGKLT 242
 DB 252 QVDTNVSEGD 262

RESULT 31

AAU35553 ID AAU35553 standard; Protein; 390 AA.

XX AC AAU35553;

XX DT 14-FEB-2002 (first entry)

XX DE Haemophilus influenzae cellular proliferation protein #194.

XX KW Antisense; prokaryotic cellular proliferation protein;

XX KW antibiotic; antibacterial; drug design.

XX OS Haemophilus influenzae.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207272P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 XX Yamamoto RT, Xu HH;
 XX MPI; 2001-611495/70.
 XX DR N-PSDB; AAS53412.
 XX PT New polynucleotides for the identification and development of
 antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11146; 511p; English.

XX The invention relates to antisense inhibitors of genes essential to
 XX prokaryotic cellular proliferation, their use in identifying the
 XX genes, their use in the discovery of novel antibiotics, the essential
 XX genes themselves and the encoded proteins. The prokaryotes used are
 XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 XX invention is also useful for the identification of potential new targets
 XX for antibiotic development. The antisense nucleic acids can also be used
 XX to identify proteins used in proliferation, to express these proteins,
 XX and to obtain antibodies capable of binding to the expressed proteins.
 XX The proteins can be used to screen compounds in rational drug discovery
 XX programmes. The antisense nucleic acid sequence is also useful to screen
 XX for homologous nucleic acids which are required for cell proliferation in
 XX a wide variety of organisms. The present sequence represents an
 XX essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 390 AA;

Query Match 7.8%; Score 156.5; DB 22; Length 390;
 Best Local Similarity 21.6%; Pred. No. 3.2e-05;
 Matches 71; Conservative 51; Mismatches 115; Indels 91; Gaps 10;

QY 75 VRAQVGGITQKLFQESGVYRAGOPLYQIDSSYEANLESARQALTA----- 122
 DB 64 VSSQVAGNVAKINADNDKVAAGDILVELDNTAKLSFEQAKSNLANAVRVEQLGFTVQ 123
 QY 123 -----QATLAKADADLARVYKPLVAEAVSRQEDVAVTAKRSNAGVKAQ----- 168
 DB 124 QLGSAVHANETSLAQAGCNLRVQLEMGAIDEXEFOHAKAEVLEAKANLNASKNQLAA 183
 QY 169 -----AAIKSAGINLRSRITAPISGFIGOSKVSSEGLTNAGDPTT 208
 DB 184 NQALLRNVPLEQPOIQINAINSLQAWMLNDRITRISPDGVYARNRVQGVAVSGAL 243
 QY 209 VLATIRQTNPMYVNTOSASEVMKLRQI-----AEGKTLADGVYIAGIFDGTVPPE 263
 DB 244 MAVVSNEMQMLEANFEKETQLTNRIGOPVKIHFDLYGNKEKDFDVIN-GIEMGTGNAF-- 300
 QY 264 KGRLLFPDVPVNESTGQITLAAVNDQNILMPLGVYRVLMQYAVNDNAFVVPQO----- 318
 DB 301 --SL--PSQNAATGNMVKVQVRV-----VRIKLD-----POQFTETP 334
 QY 319 -AVTRGAKDTVMIVNAQGM-----EPR 340
 DB 335 LRIGLSATKVRISDSSGAMLRKTEPR 362

RESULT 32

AAE30466 ID AAE30466 standard; Protein; 390 AA.

XX AC AAE30466;

XX DT 24-FEB-2003 (first entry)

XX DE H. influenzae emrA protein (multidrug resistance protein A).

XX KW Virulence; veterinary; infection; pneumonia; antimicrobial drug; vaccine;
 XX antibiotic; gene therapy; antibacterial; multidrug resistance protein A;
 XX emrA protein.

XX OS Haemophilus influenzae.

XX PN WO200277020-A2.

XX PD 03-OCT-2002.

AC ABP54530;
 XX
 DT 24-JAN-2003 (first entry)
 XX
 DE Actinobacillus pleuropneumoniae apvd protein SEQ ID NO:131.
 XX
 DE Actinobacillus pleuropneumoniae apvd protein SEQ ID NO:131.
 XX
 KW Antibacterial; vaccine; gram negative bacterial virulence gene;
 KM identification; virulence; Pasteurellaceae.
 XX
 OS Actinobacillus pleuropneumoniae.
 XX
 PN WO200275507-A2.
 XX
 PD 26-SEP-2002.
 XX
 PF 17-JAN-2002; 2002WO-US01971.
 XX
 PR 15-MAR-2001; 2001US-0809665.
 XX
 PR (PHAA) PHARMACIA & UPJOHN CO.
 PA
 PI Lowery DE, Fuller TE, Kennedy MJ;
 XX
 DR WPI, 2002-740868/80.
 DR N-PSDB; AB083530.
 XX
 PT New mutant gram-negative bacteria, useful as vaccines and for
 PT identifying new anti-bacterial agents that target virulence genes and
 PT their products -
 XX
 PS Claim 36; Page 296-297; 350pp; English.
 XX
 CC The present invention describes a gram-negative bacteria comprising a
 CC mutation in a gene, where the mutation results in decreased activity of
 CC a gene product encoded by the mutated gene. Also described is a method
 CC for producing a gram-negative bacteria mutant or an attenuated
 CC Pasteurellaceae bacteria. The mutated genes have antibacterial activity
 CC and can be used in vaccines. The gram-negative bacteria or the
 CC attenuated Pasteurellaceae bacteria can be used as vaccines in the
 CC fields of human medicine or veterinary medicine, and for identifying
 CC new antibacterial agents that target the virulence genes and their
 CC products. AB083458 to AB083578 and ABP54473 to ABP54551 represents
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 175 AA;
 Query Match 8.1%; Score 164.5; DB 23; Length 175;
 Best Local Similarity 28.2%; Pred. No. 2e-06; Indels 33; Gaps 4;
 Matches 53; Conservative 30; Mismatches 72;
 Oy 60 TVELPGRLESLRTADVRAOVGGIIQKRLFOEGSYVRAGOPLYQIDSSTYEANTLESARAO 119
 Db 5 TVIASGTIQATEQVDIGAVSGQIHITVQESQKVKKELLAVIDPRLAFTELKIAKXEL 64
 Oy 120 ATAQA-----TLAKADADLARYKPIVAEAVSROEYDAAVTAKSAAGVKAQ----- 168
 Db 65 ANASANLDTKTKINKQLQSDWBRHRLRTATNATQKETEBAKSRNTAKAEIQINQNL 124
 Oy 169 ---AAIKSAGINLNRSRITAPISGFIGOSKVSSEGTILNAGDTTYLATIRQTNPMVNVNTQ 225
 Db 125 IAKIRVEKAEETELGYTEIRSPD-----DATVISVEAOMQOTLV-TTQ 165
 Oy 226 SASFVVKL 233
 Db 166 QVPVLMKL 173
 RESULT 28
 AAU36320
 ID AAU36320 standard; Protein; 285 AA.
 XX
 AC AAU36320;
 XX

DT 14-FEB-2002 (first entry)
 XX
 DE Pseudomonas aeruginosa cellular proliferation protein #310.
 XX
 DE Pseudomonas aeruginosa cellular proliferation protein;
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyekind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI, 2001-611495/70.
 DR N-PSDB; AAS54179.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 11913; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acid can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 285 AA;
 Query Match 8.0%; Score 161.5; DB 22; Length 285;
 Best Local Similarity 26.0%; Pred. No. 7.3e-06;
 Matches 69; Conservative 40; Mismatches 101; Indels 55; Gaps 10;
 Oy 71 RTADVRA-----QVGIIQKRLFOEGSYVRAGOPLYQIDSSTYEANTLESARAO 123
 Db 35 RDAVRADVAVVADVSGWGTDLKQNVQKVDVLRIDQERYQANLEBARVAEIRH 94
 Oy 124 ATLAKADADLARYKPL-----VAEAVSROEYDAAVTAKSAAGVKAQAIAIKSAGINLN 179
 Db 95 QQVYLQNEAARSRRLGIAISAEDKEMAQINMAI-----ARSEYQELAQVKIAEINLX 149
 Oy 180 RSRITAPISGFIGOSKVSSEGTILNAGDTTYLATIRQTN---PMVNVNTQASAEVVKLRRO 236
 Db 150 RSELRARNGOVNTNLRLOGVYATAGO-AVMALVDOOSFVVAAYFEETKJLGIRVGMRAQ 208

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.

XX 30-MAY-2000; 2000US-206132P.

XX 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABN20024.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders

PS Disclosure; SEQ ID 8526; 1037pp; English.

XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 89 AA;

Query Match 8.2%; Score 165.5; DB 23; Length 89;

Best Local Similarity 43.0%; Pred. No. 6e-07; Mismatches 1; Gaps 1;

Matches 34; Conservative 17; Indels 1; Gaps 1;

QY 211 ATIRQTNPVYVNTOSASBVMKLRQIAEGKLLADGVIAVGIKFDGTVPEKRLFLPA 270

DB 1 ASVQGLDPIYVDLTQSSNDFQTKQAIEGGS-LTKDGSQVNIIMDTGQTYGHTKLEFP 59

QY 271 DPVNVESGQITLRAAVRN 289

DB 60 DVTVDSTGSGVTLRAVFPN 78

RESULT 26

ID AAB44578

XX AAB44578 standard; Protein; 175 AA.

XX AAB44578;

XX 08-FEB-2001 (first entry)

XX Virulence gene protein #58.

XX Virulence gene; antibacterial; vaccine; bacterial infection;
 KW septicemia; bronchopneumonia; rhinitis; wound infection.

XX Actinobacillus pleuropneumoniae.

XX WO200061724-A2.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US09218.

XX 09-APR-1999; 99US-0128689.

XX 10-SEP-1999; 99US-0153453.

XX (PHAA) PHARMACIA & UPJOHN INC.

XX Lowery DE, Fuller TE, Kennedy MJ;

XX WPI; 2000-647422/62.

XX N-PSDB; AAC79653.

XX Attenuated Pasteurellaceae bacteria comprising mutations in virulence
 PT genes, useful as a live attenuated vaccine against bacterial infections

PT Claim 39; Page 290; 322pp; English.

XX The family Pasteurellaceae encompasses several pathogens that infect a
 CC wide variety of animals. The present invention relates to virulence genes
 CC from Pasteurellaceae. The present sequence is a protein encoded by one
 CC such virulence gene. The virulence genes of the present invention may be
 CC mutated in order to produce an inactive gene. The inactive virulence gene
 CC may in turn be used to produce a vaccine, which is useful for treating
 CC bacterial infections such as septicemias, bronchopneumonias, rhinitis and
 CC wound infections.

XX Sequence 175 AA;

Query Match 8.1%; Score 164.5; DB 21; Length 175;

Best Local Similarity 28.2%; Pred. No. 2e-06; Mismatches 72; Indels 33; Gaps 4;

Matches 53; Conservative 30;

QY 60 TVELEGRLESIRTDVRAQVGIIQKLPQEGSVYRAGQPLVQIDSTYENLESARQL 119

DB 5 TVIASGLQATEQVDIGAVSGQIKHIVQSGVKKGELLAVIDPRLAETELKAKAEL 64

QY 120 ATAAQ-----TLAKADADLARVKPLVAARAVSROEYDAVTKRASAAGVKAQ----- 168

DB 65 AMASANLDTKIKINIKOLOSDWERHORLRIRNATSOKEETKSRKLNITKAKELQIAQNND 124

QY 169 ---AIAKSAGINLNRSRITAPISGFIGOSKVSSEGLNAGDTYVATIRQTNPMYVNTQ 225

DB 125 IAKRIVEKAETELQVTEIRSP-----DATVSVPAQNGQTLV-TTQ 165

QY 226 SASSEVMKL 233

DB 166 QVPVLMKL 173

RESULT 27

ABP54530
 ID ABP54530 standard; Protein; 175 AA.

prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 299 AA;

Query Match 8.4%; Score 169; DB 22; Length 299;
Best Local Similarity 25.4%; Pred. No. 1.7e-06;
Matches 71; Conservative 43; Mismatches 105; Indels 60; Gaps 9;

52 VHPQTALVELPGRLESLRTADVRAQVGIIQKELFQEGSYVRAQPLVYQIDSTYEAM 111

54 IRAEVSITPQVSGHVELIMD-----NQVNMGDLLITDKTPFQA 97

112 LESAPQLATQAOTLAKADADIALRYKPLVAAVSRQEVDAAVTKRSAGCVKAQAQAI 171

98 ELNAAQALAKQOSDLAKANNENRRRL-SQNFISAEELDTANLVKAMQASVDAQAQTL 156

172 KSAGINLRSRTAPISGRIGSKVSEGLNAGTTLATIRQTNPMVNVNTQASAEVM 231

157 KQAGQQLQTEIRAPVSGWVNLTRIGVADTG-----KPLFALVDSISFYI 205

232 -----KLRFQIAEKG-----LTAADG-----VIAVGKFDGDTVYPEKGRLLFADPVN 275

206 GFEEETKL-RHIREGAPAQITLYSDNKTLCQHSVSGIRAIYQSVESDSSL-----PVK 260

276 ESTGQITLRAAVPNDQNTLMPELVYRVLMDOVAVDNAAFV 314

261 PNVFWRLAQRVP-----VRFALDKVPGDITLV 288

RESULT 24

AAU38257 standard; Protein; 309 AA.

AAU38257;

14-FEB-2002 (first entry)

Salmonella typhi cellular proliferation protein #148.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Salmonella typhi.

MO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI; 2001-611495/70.

N-PsDB; AAS56116.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

Example 3; Seq ID No 13850; 511pp; English.

The invention relates to antisense inhibitors of genes essential to

prokaryotic cellular proliferation, their use in identifying the

genes, their use in the discovery of novel antibiotics, the essential

genes themselves and the encoded proteins. The prokaryotes used are

Escherichia coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

pneumoniae, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The

invention is also useful for the identification of potential new targets

for antibiotic development. The antisense nucleic acids can also be used

to identify proteins used in proliferation, to express these proteins,

and to obtain antibodies capable of binding to the expressed proteins.

The proteins can be used to screen compounds in rational drug discovery

programmes. The antisense nucleic acid sequence is also useful to screen

for homologous nucleic acids which are required for cell proliferation in

a wide variety of organisms. The present sequence represents an

essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part

of the printed specification, but was obtained in electronic

format directly from WIPO at

ftp.wipo.int/pub/published_pct_sequences.

Sequence 309 AA;

Query Match 8.4%; Score 169; DB 22; Length 309;

Best Local Similarity 24.5%; Pred. No. 1.8e-06;

Matches 63; Conservative 45; Mismatches 103; Indels 46; Gaps 7;

75 VRAQVGITIQRLFQEGSYVRAQPLVYQIDSTYEAMLESAPQLATQAOTLAKA--DAD 132

70 VTPQVSGSITQNLKNDQFNVAAGDVLFPVIDTPPHIELNQAQDLAKQOSDLAAANNAD 129

133 LARYKPLVAAVSRQEVDAAVTKRSAGCVKAQAQAIKSAGINLRSRTAPISGRIG 192

130 RRRH---LSRNYIAEEDDSANLVKAMQASVDVALTLKQAGQOLSGTEVKAPVSGWVT 186

193 QSKVSEGLTNAGDVTTLATIRQTNPMVNVNTQASAEVM-----KLRR-----QI 237

187 NLSTRGTGYASTG-----KPLFALVDSISFYMGVFEEETKLRIHREGPALITL 235

238 AEGKILAADGVIAVGKFDGDTVYPEKGRLLFADPVNVESTGQITLRAAVPNDQNTLM 297

236 YSGWVKLQGHGSGIRAIYDQSVESDGLV-----PDIKPNPWRRLAQRVP----- 282

298 LYRVLMDOVAVDNAAFV 314

283 --VRIEFDALPDQDITLV 297

RESULT 25

ABP04272

ID ABP04272 standard; Protein; 89 AA.

ABP04272;

24-JUN-2002 (first entry)

Human ORF protein sequence SEQ ID NO:8526.

Human; open reading frame; ORF; gene therapy; cancer; cirrhosis;

QY 254 KPDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDONILMPGLYRVLMDOVAVDNAF 313
DB 299 GFSAGT-----GSASLLIPDAQNAIGNWKVORP-----VAIVLRDVD--- 339
QY 314 VPPQAVTRGAKDIYVIVNAOGGMEPRE---VTAQOOGTNM 352
DB 340 ---RHPLRIGLSMTYKVDTSAAAGAVSKTPGALPEMESTW 378
RESULT 22
AAG3061
ID AAG3061 standard; Protein; 659 AA.
XX AAG3061;
AC
XX 26-SEP-2001 (first entry)
DT
XX C glutamicum protein fragment SEQ ID NO: 6815.
DE
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX EPI108790-A2.
PN
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-0127688.
PF
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH68280.
XX
XX Novel polynucleotides derived from Corynebacterium, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 6815; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 659 AA;
Query Match 8.7%; Score 175; DB 22; Length 659;
Best Local Similarity 22.1%; Pred. No. 1.6e-06;
Matches 102; Conservative 81; Mismatches 194; Indels 84; Gaps 16;
QY 5 AFKAMPAAALAAVALVIVSSCGKGDAAAGGAPAGREAPAVGVVTHPQTVALTVEL- 63
DB 223 ALKAREQADAAEIERLRADPGLNN-----DRSNINDVIGLDERESIASSEELA 274

QY 64 ----PGRLESLRTADVAVOAGGIIIOKRLFOGGSYVRAGOPLYOIDSSTYEANLESAPQL 119
DB 275 QARAGGLEAVAAAE--AKVAGLEQSIASKSTSTPSPDQTYLQ---SYTLEAEARRVA 328
QY 120 ATAQA-----TLAKADADLARYKPLVAEAVSRQEDAAV---TAKSAEAGVKA 166
DB 329 STTLEAIEAERIVYDSIGKVDSELAQAQAV-ABEHSAAQ-DAALGLETAQLSTOHOLEA 386
QY 167 AQAARKSA-----GININRSRITAPISGFI-----GQSKVEBGTLLN 203
DB 387 QSSAIDALGLAVDNEAATRSQSLRMDINNTTVRSPPSGIVSSVOAAQOPAPAGALLS 446
QY 204 AGDTT---VLATIRQTPMYVNV-----TQSAEWMKLRRQIAEGKLLADGVYAVGIRK 254
DB 447 VADSELEKITHANVEAISNVTIGSRFTFTTPTSGTGFAGRVSKVPIAA----- 497
QY 255 FDDGTVYPEKGRLLFADPVNVESTGQITLRAAVNDONILMPGLYRVLMDOVAVDNAFV 314
DB 498 ---AASAPATGEGAAAGATTNTDVTPEIEISVTGDRGLNLGGSARVRIVHEIAPHVLT 554
QY 315 VPPQAVTRG--AKDTVMIVNAOGGMEPREVTVAAQOOGTNMIVT--SGLKDGKRVVEGISI 371
DB 555 VPLEAVYKNDGKDAVLIISDNNKVEEVEKTAESDDFDIAVSGAGISEDAVRLTQPGNY 614
QY 372 AGITGAKKVTPEKWAASENQAAPQSGVQTASEAKTASEAE 412
DB 615 RGLTGE---TVKLHADTVEQALAPFSPAPAPPDPAAPVSAK 652
RESULT 23
AAU34575
ID AAU34575 standard; Protein; 299 AA.
XX AAU34575;
AC
XX
XX 14-FEB-2002 (first entry)
DT
XX
XX E. coli cellular proliferation protein #156.
DE
XX
XX Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
XX Escherichia coli.
OS
XX WO200170955-A2.
PN
XX 27-SEP-2001.
PD
XX 21-MAR-2001; 2001WO-US09180.
PF
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (EIT-) EITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52434.
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10168; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to

KM cardiovascular disease; immune system disorder; organ transplantation;
 KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KM hypothyroidism; cholesterol ester storage disease; infection; vlnnerary;
 KM vasoprotective; antipsoriatic; antidiabetic; cyclostatic; nootropic;
 KM neuroprotective; antithrombotic; anticoagulant; thrombolytic;
 KM cardiac; hypotensive; antihypertoid; antiinflammatory; immunomodulator;
 KM dermatological; analgesic; virucide; antibacterial; fungicide.
 XX Homo sapiens.
 OS
 XX WO200190366-A2.
 PN
 XX 29-NOV-2001.
 PD
 XX 24-MAY-2001, 2001WO-US17076.
 PF
 XX 24-MAY-2000; 2000US-206690P.
 PR
 XX 24-MAY-2000; 2000US-206690P.
 PA (CURA-) CURAGEN CORP.
 XX Leach MD, Shimkets RA;
 P1
 XX WPI; 2002-106200/14.
 DR N-PSDB; ABN75407.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation -
 PS
 XX Claim 10; Page 445; 2508pp; English.
 XX
 CC Sequences ABP1028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC proteins, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 CC
 XX
 SQ Sequence 102 AA;

Query Match 10.7%; Score 217; DB 23; Length 102;
 Best Local Similarity 40.2%; Pred. No. 2.1e-11;
 Matches 41; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

QY 221 VNVTSASEVWKLKRLQIAEGKLLAADGVIANVGIKFDGTVPEKGRLLFADPVNVESTGQ 280
 Db 1 VDVTSNSDFRLKQELSGALKOENGKAKVKKLLLENQAYAOEGTLEFSDVTDETGS 60
 QY 281 ITRAAVNDONIMPGLYRVLDQVAVDAFVVPQOAVTR 322
 Db 61 ITRAVFPNPMDTLPGHFFVARLDEGVRSALLVPOGAVTR 102
 RESULT 21
 ABP79084
 ID. ABP79084 standard; Protein, 394 AA.
 XX
 AC ABP79084;
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae amino acid sequence SEQ ID 4698.
 XX
 KM Antibacterial; infection; vaccine; gene therapy.
 KM
 OS Neisseria gonorrhoeae.
 XX
 PN WO200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-1B02069.
 XX
 PR 12-FEB-2001; 2001GB-0003424.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizsa M, Masignani V, Monaci E;
 XX
 DR WPI; 2003-058415/05.
 DR N-PSDB; ABZ40054.
 XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection -
 XX
 PS Disclosure; Page 533; 815pp; English.
 XX
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 CC
 XX
 SQ Sequence 394 AA;
 Query Match 9.3%; Score 188; DB 24; Length 394;
 Best Local Similarity 24.3%; Pred. No. 5.3e-08;
 Matches 83; Conservative 42; Mismatches 127; Indels 90; Gaps 10;
 QY 71 RTAVRQVGGITIKRLPQESGYRAGOPLYQIDSS---TYE----- 109
 Db 67 RVQVTPQKGGTVKRVLHDDTDVAVKGDVLAIVDDNDVLAIEAKNELVOAVRONRRON 126
 QY 110 ANLESARQQLATAQATLAKADADLARVYKPLVAEAASVROEYDAVTAKRSAEAGVKAQA 169
 Db 127 AATQAGAGVALLRADARARQDDLRSSALAEASVAAEELAHARTAVSQQAIVKALALA 186
 QY 170 -----AIKSGAGINLNSRTAPISGIFGSKVSEGTLLN 203
 Db 187 EESSARAALGGDVSLRQPEVQTAIGRLKQAMINLRTQVRAPADGVAKRSVQGGQVA 246
 QY 204 AG-----DTTVALATRTQNPVNVNTQASAEVWKRRRQIAEKKLLAADGVIANVGI 253
 Db 247 AGAPLMAVPLSDVWVADANFKETOLRMKTIQOPAEIVSDI-----YKQIYVRGRVA--- 298

CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.

XX Sequence 334 AA;

Query Match 13.4%; Score 271; DB 20; Length 334;

Best Local Similarity 28.6%; Pred. No. 1.9e-15;

Matches 94; Conservative 56; Mismatches 135; Indels 44; Gaps 9;

QY VRAVGGIIORLRFEGSVYRAGQPLVQIDS-----STVEANLESARQATATQAATLA 127

Db 3 VQAQASGQIKILVYKLGQOVKKGDLIAEINSTQNTLNTKSKLETYQAKLVSAQIALG 62

QY 128 KADADLARYKPLVAEAVSROEYDAVTAKRSAEAGVKAQAIAKSAGINLNR----- 180

Db 63 SAEKKYRKQALMKENATSKEDLESADAFAPAAKANVALEKALIRKSKISINTASELGY 122

QY 181 SRTIAPISGFIQSGKVSSEGLTLNAGDTTVLATIRQTNPMYVNTQSASEVKKLRQIAEG 240

Db 123 TRITATMDGTVAALIVEEGQTVNAA-----QSTPTIYQLAN--LDMMLNKQIAEG 171

QY 241 KL--LAADGVAVGIRKDDGTVVPEKGRLEFADPVV-----NESTGQIT-----LR 284

Db 172 DITVKRAGQDISFTILSEPT--PIKALDSVDPGLTTMSSGGYNSSTDITASNAVYYAR 229

QY 285 AAVPNDQNIIMPGLYVRLMDQVAVDNAFVVPQOAV--TRGAKDTVMIVNAOGMEPREVT 343

Db 230 SFVNPDPGLKATGMTTQNTVEIDGKVNLIIPSLTVKNRGKAFVRLVAGADGKAERIR 289

QY 344 VAAQOGTNWIVTSGLKDGDKRVVVGISIA 372

Db 290 TGMKDSMNTVEVSGLKEGDKVVISIETAA 318

Db

RESULT 19

AAV74870 ID AAV74870 standard; Protein; 334 AA.

XX AC AAV74870;

XX DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 290 protein sequence SEQ ID NO:1214.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

XX antibiotic; gene therapy.

XX Neisseria meningitidis.

XX OS Neisseria meningitidis.

XX PN MO9957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 09-OCT-1998; 98US-0103796.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

XX

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalati E, Scarselli M;
PI Tetteijn H, Venter JC;
XX WPI; 2000-062150/05.
DR N-PSDB; AA253632.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics

PS Claim 2; Page 677; 1453p; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941

XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

XX and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent

XX PCR primers used in the exemplification of the present invention. The

XX polypeptides, the polynucleotides, antibodies and compositions of

XX the invention can be used as vaccines, as diagnostic reagents, and as

XX immunogenic compositions. The polypeptides can be used in the

XX manufacture of medicaments for treating or preventing infection due to

XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the

XX presence of Neisseria bacteria, or to raise antibodies. They may also

XX be used to screen for agonists or antagonists, which may themselves

XX have use as antibacterial agents. The polynucleotides of the invention

XX may also be used in gene therapy protocols.

XX Sequence 334 AA;

Query Match 13.4%; Score 271; DB 21; Length 334;

Best Local Similarity 28.6%; Pred. No. 1.9e-15;

Matches 94; Conservative 56; Mismatches 135; Indels 44; Gaps 9;

QY VRAVGGIIORLRFEGSVYRAGQPLVQIDS-----STVEANLESARQATATQAATLA 127

Db 3 VQAQASGQIKILVYKLGQOVKKGDLIAEINSTQNTLNTKSKLETYQAKLVSAQIALG 62

QY 128 KADADLARYKPLVAEAVSROEYDAVTAKRSAEAGVKAQAIAKSAGINLNR----- 180

Db 63 SAEKKYRKQALMKENATSKEDLESADAFAPAAKANVALEKALIRKSKISINTASELGY 122

QY 181 SRTIAPISGFIQSGKVSSEGLTLNAGDTTVLATIRQTNPMYVNTQSASEVKKLRQIAEG 240

Db 123 TRITATMDGTVAALIVEEGQTVNAA-----QSTPTIYQLAN--LDMMLNKQIAEG 171

QY 241 KL--LAADGVAVGIRKDDGTVVPEKGRLEFADPVV-----NESTGQIT-----LR 284

Db 172 DITVKRAGQDISFTILSEPT--PIKALDSVDPGLTTMSSGGYNSSTDITASNAVYYAR 229

QY 285 AAVPNDQNIIMPGLYVRLMDQVAVDNAFVVPQOAV--TRGAKDTVMIVNAOGMEPREVT 343

Db 230 SFVNPDPGLKATGMTTQNTVEIDGKVNLIIPSLTVKNRGKAFVRLVAGADGKAERIR 289

QY 344 VAAQOGTNWIVTSGLKDGDKRVVVGISIA 372

Db 290 TGMKDSMNTVEVSGLKEGDKVVISIETAA 318

Db

RESULT 20

ABP31381

ID ABP31381 standard; Protein; 102 AA.

XX AC ABP31381;

XX DT 08-JUL-2002 (first entry)

XX Human glycoprotein-like ORF354 protein, SEQ ID NO:708.

XX Human; ORF; open reading frame; ORF; drug screening; diagnosis;

XX disease monitoring; cytokine; cell proliferation; cell differentiation;

XX immune modulation; haematopoiesis regulation; tissue growth;

XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;

XX chromolytic; tumour inhibition; bodily characteristic; fertility;

XX behaviour; cancer; proliferative disorder; neurological disorder;

XX

```

QY 120 ATAQATLAKADADLARYKPLVAEAEVSROEYDAVTAKRSAAGYKAAQAIAKSAGINLN 179
Db 113 VSAQIALGSAEKKYKROALWMDATAKEDLESADALAAKAVAELEKALIROSKISIN 172
QY 180 R-----SRITAPISGFIGSKVSEGLTLNAGDTTVALTQINPMYVNTQSASEVWK 232
Db 173 TAESLGTTRITATMDGTVALVLEEGQTVA-----QSPPTIVQLAN--LDMWL 221
QY 233 LRQIAEGL--LAADGVIAVGIRKFDGTVYPEKGRLLPADPV-----NESTGQIT 282
Db 222 NKMQIAEGDITKVKAGODISFTILSEPT--PIKAKLDSVDPGLTMSGGYNSSTDAS 279
QY 283 -----LRAAVPNQNTILMPGLYVRVLDQVAVDNAPVPOQAV--TRGAKDTVMVINAOG 335
Db 280 NAVYYAASFVNPDPGKLATGTTONTVEIDGVKNVLIIPSLTVKNRGGRAPRVVLGADG 339
QY 336 GMEPEVTVAAOQGTNMIYVTSGLKXGDKVVEGIGSIA 372
Db 340 KAAEREIRTGMRDSMNTKVGSLKEGDKVVISSEITPA 376

RESULT 17
AAU73013
ID AAU73013 standard; Protein; 392 AA.
AC AAU73013;
XX
XX 12-MAR-2002 (first entry)
DT
DE Neisseria meningitidis virulence protein #103.
XX
XX Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
XX infection; Gram-negative bacteria; antimicrobial.
XX
XX Neisseria meningitidis.
OS
XX WO200185772-A2.
XX
XX 15-NOV-2001.
PD
XX
XX 08-MAY-2001; 2001MO-GB02003.
PR
XX 08-MAY-2000; 2000GB-0011108.
XX
XX (MICR-) MICROSCIENCE LTD.
PA
XX
XX Tang C.
PI
XX WPI; 2002-066593/09.
DR
XX N-PSDB; AAS97298.
XX
XX New peptide encoded by operon including virulence genes of Neisseria
XX meningitidis, useful as vaccine component for treating or preventing
XX meningitis and for identifying antimicrobial drug
PT
XX
XX Claim 4; Page 419-420; 423pp; English.
XX
XX The invention relates to a peptide (I) encoded by an operon (II) of
XX Neisseria meningitidis including virulence genes, or a related molecule
XX having a 40% sequence similarity at the peptide or nucleotide level in a
XX Gram-negative bacterium, or its functional fragment, for therapeutic or
XX diagnostic use. (I) and (II) are useful in the manufacture of a
XX medicament for treating or preventing a condition (e.g., meningitis)
XX associated with infection by Neisseria or Gram-negative bacteria. The
XX product is useful for veterinary treatment and in a screening assay for
XX the identification of an antimicrobial drug. The vaccines have
XX prophylactic applications. AAU72911-AAU73014 represent N. meningitidis
XX virulence proteins of the invention.
XX
XX Sequence 392 AA.
SQ

```

```

Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;
QY 7 KAMRAALAAVALVLSGCGKGDAAOGGAPAGRAPAVGVVTHPQVALTYELPGR 66
Db 3 KMMKWAAVAVALA-----AVMGSGEYLKREPOAAVITETVRGDISRTVSATGE 52
QY 67 LESLRTADVRAQVGGIIOKRLFOEGSYVRAQOPLYQIDS-----STYEANLESARQOL 119
Db 53 ISPSNLVSVAQASQIKLVKLGQOVKKDLAEINSTQNTLNTKESKLETYQAKL 112
QY 120 ATAQATLAKADADLARYKPLVAEAEVSROEYDAVTAKRSAAGYKAAQAIAKSAGINLN 179
Db 113 VSAQIALGSAEKKYKROALWMDATAKEDLESADALAAKAVAELEKALIROSKISIN 172
QY 180 R-----SRITAPISGFIGSKVSEGLTLNAGDTTVALTQINPMYVNTQSASEVWK 232
Db 173 TAESLGTTRITATMDGTVALVLEEGQTVA-----QSPPTIVQLAN--LDMWL 221
QY 233 LRQIAEGL--LAADGVIAVGIRKFDGTVYPEKGRLLPADPV-----NESTGQIT 282
Db 222 NKMQIAEGDITKVKAGODISFTILSEPT--PIKAKLDSVDPGLTMSGGYNSSTDAS 279
QY 283 -----LRAAVPNQNTILMPGLYVRVLDQVAVDNAPVPOQAV--TRGAKDTVMVINAOG 335
Db 280 NAVYYAASFVNPDPGKLATGTTONTVEIDGVKNVLIIPSLTVKNRGGRAPRVVLGADG 339
QY 336 GMEPEVTVAAOQGTNMIYVTSGLKXGDKVVEGIGSIA 372
Db 340 KAAEREIRTGMRDSMNTKVGSLKEGDKVVISSEITPA 376

RESULT 18
AAV38882
ID AAV38882 standard; Protein; 334 AA.
AC AAV38882;
XX
XX 08-OCT-1999 (first entry)
DT
DE Neisseria meningitidis antigen encoded by a partial ORF85.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX
XX Neisseria meningitidis.
OS
XX WO9924578-A2.
XX
XX 20-MAY-1999.
PD
XX
XX 09-OCT-1998; 98WO-1B01665.
PR
XX 01-SEP-1998; 98GB-0019016.
PR
XX 06-NOV-1997; 97GB-0023516.
PR
XX 14-NOV-1997; 97GB-0024190.
PR
XX 18-NOV-1997; 97GB-0024386.
PR
XX 27-NOV-1997; 97GB-0025158.
PR
XX 10-DEC-1997; 97GB-0026147.
PR
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Mesigiani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
DR
XX N-PSDB; AA212304.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
PT
XX
XX Claim 4; Page 427; 524pp; English.
XX
XX Amino acid sequences AAV38499-Y38944 represent Neisseria meningitidis
XX
XX

```

Query Match 15.3%; Score 308; DB 23; Length 392;
 Best Local Similarity 27.5%; Pred. No. 1.3e-18;

antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
antibacterial; gene therapy.

Neisseria meningitidis.

MO9957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.
31-JUL-1998; 98US-0094869.
02-SEP-1998; 98US-0098994.
02-SEP-1998; 98US-0099062.
09-OCT-1998; 98US-0103749.
09-OCT-1998; 98US-0103794.
09-OCT-1998; 98US-0103796.
25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.
(GENO-) INST GENOMIC RES.

Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
Peterson J, Pizzi M, Rapunoli R, Ratti G, Scalato E, Scarselli M;
Tettelin H, Venter JC;

WPI: 2000-062150/05.
N-PSDB; AA253633.

Novel Neisserial polypeptides predicted to be useful antigens for
vaccines and diagnostics

Claim 2; Page 678; 1453pp; English.

AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenetic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisserial bacteria (e.g. meningitis and septicemia), to detect the
presence of Neisseria bacteria, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention
may also be used in gene therapy protocols.

Sequence 392 AA;

Query Match 15.3%; Score 308; DB 21; Length 392;
Best Local Similarity 27.5%; Pred. No. 1.3e-18;
Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;

7 KAMRAAALAAVALVLSGCGDAAGGQAPGAPAPVGVTVHPQTVALTYELPGR 66
3 KMMKAAVAVAALVLSGCGDAAGGQAPGAPAPVGVTVHPQTVALTYELPGR 52
AVWGMSTLKEPQAAVITETVRGDISRTYSATGE

67 LESLRTADVAVGIIIOKRLFOGSGYVRAGQPLVQIDS-----STYEANLESARQ 119
53 ISPSNLVSVGAQASQIKLVKLGQOVKKGDILAEINSTQNTLNTEKSKLETYQAKL 112

120 ATAQATLAKADADLARKYPLVAEAVSRQETPAATYAKRSAAAGYKAAQAAKSGINLN 179
113 VSAQIALGSAEKYKRAALWMDATKEDLESADALAAAKANVAELKALIRQKISIN 172
180 R-----SRIAPISGFIGSKVSEGLTLNAGDTTVALTIRQTPMVVNTVQASSEWK 232
173 TASELGTIRITATMDGYVALVEGGTVNAA-----QSTPTTVQLAN--LDMML 221
233 LRRQIAEGKL--LAADGVIAGIKFDDGTVPYKGRLLFADPVP-----NESTGOIT 282

222 NKMOIAEGDITKVAQODISFTLISEPDT--PIKAKLSDVDPGLTTMSSGGYNSSTDTAS 279
283 -----LRAAVPNDQNLIMPGLVYVLMDDQVAVDNAAFVPOQAV-TRGAKDTWIMVNAQG 335
280 NAVTYVARSFVPNDGLATGTTQNTVEIDGVKAVLII8LTYKANRGRAFVVLGADG 339
336 GMEPEVTVAAQOQCTNWIIVTSGLDGDKVVEGISA 372
340 KAAREIRTGMRDSEMTVEVSGLKEGDKVYISEITAA 376

RESULT 16
AAU72963
ID AAU72963 standard; Protein; 392 AA.
XX AA
XX AAU72963;
XX AC
XX 12-MAR-2002 (first entry)
DT XX
XX Neisseria meningitidis virulence protein #53.
DE XX
XX Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
KM infection; Gram-negative bacteria; antimicrobial.
XX Neisseria meningitidis.
OS
XX WO200185772-A2.
PN
XX 15-NOV-2001.
PD
XX 08-MAY-2001; 2001WO-GB02003.
PF
XX 08-MAY-2000; 2000GB-0011108.
PR
XX (MICR-) MICROSCIENCE LTD.
PA
XX Tang C;
PI
XX WPI: 2002-066593/09.
XX N-PSDB; AAS97248.
DR
XX New peptide encoded by operon including virulence genes of Neisseria
PT meningitidis, useful as vaccine component for treating or preventing
PT meningitis and for identifying antimicrobial drug

Claim 4; Page 226-228; 423pp; English.

The invention relates to a peptide (I) encoded by an operon (II) of
Neisseria meningitidis including virulence genes, or a related molecule
having a 40% sequence similarity at the peptide or nucleotide level in a
Gram-negative bacterium, or its functional fragment, for therapeutic or
diagnostic use. (I) and (II) are useful in the manufacture of a
medicament for treating or preventing a condition (e.g., meningitis)
associated with infection by Neisseria or Gram-negative bacteria. The
product is useful for veterinary treatment and in a screening assay for
the identification of an antimicrobial drug. The vaccines have
prophylactic applications. AAU72911-AAU73014 represent N. meningitidis
virulence proteins of the invention.

Sequence 392 AA;

Query Match 15.3%; Score 308; DB 23; Length 392;
Best Local Similarity 27.5%; Pred. No. 1.3e-18;
Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;

7 KAMRAAALAAVALVLSGCGDAAGGQAPGAPAPVGVTVHPQTVALTYELPGR 66
3 KMMKAAVAVAALVLSGCGDAAGGQAPGAPAPVGVTVHPQTVALTYELPGR 52
AVWGMSTLKEPQAAVITETVRGDISRTYSATGE

67 LESLRTADVAVGIIIOKRLFOGSGYVRAGQPLVQIDS-----STYEANLESARQ 119
53 ISPSNLVSVGAQASQIKLVKLGQOVKKGDILAEINSTQNTLNTEKSKLETYQAKL 112

CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhoea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

CC Sequence 392 AA;

Query Match 15.3%; Score 308; DB 20; Length 392;
 Best Local Similarity 27.7%; Pred. No. 1.3e-18;
 Matches 110; Conservative 63; Mismatches 170; Indels 54; Gaps 10;

```

OY 7 KAMRAAALAAVALVLSGCKGDAAGGAPAGREAPAPVGVTVHPQTVALTYELPCR 66
DB 3 KMMKMAAVALVAAA-----AWGGMWYLYKPEPOAAVITTAVRGDISRTVSATGE 52
OY 67 LESLRTADVRAQVGIIQKRLFOEGSYVRAGOPLYOISSTY-----EAMLESARAOL 119
DB 53 ISPSNLVSVGAQASGOIKLVKLGQGVKKGDLAEINSTTQNTIMDKSKLETYYQAKL 112
OY 120 ATAQATLAKADADLARIKPLVAEAVSRQEDAAVTAKRSAGAKAAQAIAKSGINLN 179
DB 113 VSAQIALGSAEKKYKROAALMKDDATSKEDLESADALAAKAVVAELKALIRQKISIN 172
OY 180 R-----SRTAPISGFIGOSKVS EGTILNAGDTTLATITROTNPVYNTQSASEVWK 232
DB 173 TAESDLGYTRITATMDGTVVAIPVEEGQTVNAA-----QSTPTTVQLAN--LDMML 221
OY 233 LRROIAEGKL--LAADGVAVGIRKFPDDGTVYPEKGRLLFADPVV-----NESTGOIT 282
DB 222 NKMOIAEGDITKVRAGDISFTLISEPDT--PIKAKDSVDPGGLTMSGGVNSTDTAS 279
OY 283 -----LRAAVNDQNTILMPGLVYRVLMDOVAVDNAFVVPQAAV--TRGAKDTVMITVNAOG 335
DB 280 NAVVYVARSFVNPDPGKATGMTQNTVEIDGVKNVLLIPSLTVKNGKGAFAVRLGADG 339
OY 336 GMEPREVTVAAOQGTNNIVTSGLKDGDKVYVEGISIA 372
DB 340 KAVEREIRTKGMSMTEVKSGLKEGDKVVISETIAA 376

```

RESULT 14

AA74869 standard; Protein; 392 AA.

AA74869;

21-MAR-2000 (first entry)

Neisseria gonorrhoeae ORF 290 protein sequence SEQ ID NO:1212.

Neisseria meningitidis; *Neisseria gonorrhoeae*; antigen; vaccine;
 antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 antibacterial; gene therapy.

Neisseria gonorrhoeae.

MO9957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

02-SEP-1998; 98US-0099062.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

09-OCT-1998; 98US-0103796.

25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scallato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR WPI: 2000-062150/05.
 DR N-ESDB; AA253631.

Novel *Neisseria* polypeptides predicted to be useful antigens for
 vaccines and diagnostics -
 Claim 2; page 676; 1453pp; English.

AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
 represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 PCR primers used in the exemplification of the present invention. The
 polypeptides, the polynucleotides, antibodies and compositions of
 the invention can be used as vaccines, as diagnostic reagents, and as
 immunogenic compositions. The polypeptides can be used in the
 manufacture of medicaments for treating or preventing infection due to
Neisseria bacteria (e.g. meningitis and septicaemia), to detect the
 presence of *Neisseria* bacteria, or to raise antibodies. They may also
 be used to screen for agonists or antagonists, which may themselves
 have use as antibacterial agents. The polynucleotides of the invention
 may also be used in gene therapy protocols.

Sequence 392 AA;

Query Match 15.3%; Score 308; DB 21; Length 392;
 Best Local Similarity 27.7%; Pred. No. 1.3e-18;
 Matches 110; Conservative 63; Mismatches 170; Indels 54; Gaps 10;

```

OY 7 KAMRAAALAAVALVLSGCKGDAAGGAPAGREAPAPVGVTVHPQTVALTYELPCR 66
DB 3 KMMKMAAVALVAAA-----AWGGMWYLYKPEPOAAVITTAVRGDISRTVSATGE 52
OY 67 LESLRTADVRAQVGIIQKRLFOEGSYVRAGOPLYOISSTY-----EAMLESARAOL 119
DB 53 ISPSNLVSVGAQASGOIKLVKLGQGVKKGDLAEINSTTQNTIMDKSKLETYYQAKL 112
OY 120 ATAQATLAKADADLARIKPLVAEAVSRQEDAAVTAKRSAGAKAAQAIAKSGINLN 179
DB 113 VSAQIALGSAEKKYKROAALMKDDATSKEDLESADALAAKAVVAELKALIRQKISIN 172
OY 180 R-----SRTAPISGFIGOSKVS EGTILNAGDTTLATITROTNPVYNTQSASEVWK 232
DB 173 TAESDLGYTRITATMDGTVVAIPVEEGQTVNAA-----QSTPTTVQLAN--LDMML 221
OY 233 LRROIAEGKL--LAADGVAVGIRKFPDDGTVYPEKGRLLFADPVV-----NESTGOIT 282
DB 222 NKMOIAEGDITKVRAGDISFTLISEPDT--PIKAKDSVDPGGLTMSGGVNSTDTAS 279
OY 283 -----LRAAVNDQNTILMPGLVYRVLMDOVAVDNAFVVPQAAV--TRGAKDTVMITVNAOG 335
DB 280 NAVVYVARSFVNPDPGKATGMTQNTVEIDGVKNVLLIPSLTVKNGKGAFAVRLGADG 339
OY 336 GMEPREVTVAAOQGTNNIVTSGLKDGDKVYVEGISIA 372
DB 340 KAVEREIRTKGMSMTEVKSGLKEGDKVVISETIAA 376

```

RESULT 15

AA74871 standard; Protein; 392 AA.

AA74871;

21-MAR-2000 (first entry)

Neisseria meningitidis ORF 290 protein sequence SEQ ID NO:1216.

Neisseria meningitidis; *Neisseria gonorrhoeae*; antigen; vaccine;

Neisseria meningitidis; *Neisseria gonorrhoeae*; antigen; vaccine;

Neisseria meningitidis; *Neisseria gonorrhoeae*; antigen; vaccine;

Neisseria meningitidis; *Neisseria gonorrhoeae*; antigen; vaccine;

```

Db      173 TAESDLGYTRITATMDGTVAIIPVEEGQTVA-----QSTPTIVQLAN--LDMWL 221
Oy      233 LRQIAEGKL--LAADGVIAVGKIFDGTVPPEKRLIFADPV-----NESTGQIT 282
Db      222 NKMOIAEGDITKVKAGODISFTLISEPDT--PIKAKLDSVDPGLTMSGGVNSTDTAS 279
Oy      283 -----LRAAVPNQNIIMPGLYVRVLMDOVAVDNAFVVPQOAV--TRGAKDTVMIVNAOG 335
Db      280 NAVYYIARSPFVNPDPGLATGKTONTVEIDGVKNVLIIPSLTYKRNKGRAFAVRLGADG 339
Oy      336 GMEPREVTVAOQOGTNNIVTSGLKDGDKRVVEGIGSIA 372
Db      340 KAVEREIRTKGMDSMNTEVKSGLKEGDKVVISETIAA 376

RESULT 12
AY38883
ID      AAY38883 standard; Protein; 392 AA.
XX
AC      AAY38883;
XX
DT      08-OCT-1999 (first entry)
XX
DE      Neisseria meningitidis strain A antigen encoded by ORF85.
XX
KM      Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX      treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX
OS      Neisseria meningitidis.
XX
PN      MO9924578-A2.
XX
PD      20-MAY-1999.
XX
PF      09-OCT-1998; 98WO-IB01665.
XX
PR      01-SEP-1998; 98GB-0019016.
PR      06-NOV-1997; 97GB-0023516.
PR      14-NOV-1997; 97GB-0024190.
PR      18-NOV-1997; 97GB-0024386.
PR      27-NOV-1997; 97GB-0025158.
PR      10-DEC-1997; 97GB-0026147.
PR      14-JAN-1998; 98GB-0000759.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR      MPI; 1999-327407/27.
XX      N-PSDB; AA212305.
XX
PT      Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX      diagnosis, treatment and prevention of infection
XX
PS      Claim 4; Page 428-429; 524pp; English.
XX
CC      Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX      and N. gonorrhoeae antigenic proteins. They are encoded by open
XX      reading frames (ORFs) AA21972-Z1358. The antigenic proteins,
XX      their fragments, their nucleic acids and antibodies are used for
XX      diagnosis, prevention (as vaccines) or treatment of Neisseria
XX      infections, such as meningitis, septicemia and gonorrhea. Both
XX      organisms are closely related. Fragments of the nucleic acids
XX      are useful as hybridisation probes and antisense reagents.
XX
SQ      Sequence 392 AA;

Query Match      15.3%; Score 308; DB 20; Length 392;
Best Local Similarity 27.5%; Pred. No. 1.3e-18;
Matches 109; Conservative 65; Mismatches 169; Indels 54; Gaps 10;

Oy      7 KAMRAAALAAVALVLSGCKGDAAGGQAPAGREAPAVVGVTVHPQTALVTELPGR 66

```

```

Db      3 KMMKMAAVALAA-----AVWGWSYLKPEFOAAVITETVRGDISRTVSARGE 52
Oy      67 LESIRTDVRAQVGCIIOKRLFOGGSYVRAGQPIYOIDS-----STYEANLESARQL 119
Db      53 ISPSNLVSGAQASGQIKKIVKLGQOVKKGDLIAEINISQNTLTNTERSKLETYQAKL 112
Oy      120 ATAQATLAKADADLARPKPIVAAEAVRSOEYDAVTAKRSAAGVKAQAIAIKSAGINLN 179
Db      113 VSAQIALGSAEKYKRRQALMKODATAKEDELSQODLAAAKANVAELKILIRGSKISIN 172
Oy      180 R-----SRITAPISGFQOSKVSSEGLTNAGDTVLATIRQTNPMYVNTQSASEVMK 232
Db      173 TAESDLGYTRITATMDGTVAIIPVEEGQTVA-----QSTPTIVQLAN--LDMWL 221
Oy      233 LRQIAEGKL--LAADGVIAVGKIFDGTVPPEKRLIFADPV-----NESTGQIT 282
Db      222 NKMOIAEGDITKVKAGODISFTLISEPDT--PIKAKLDSVDPGLTMSGGVNSTDTAS 279
Oy      283 -----LRAAVPNQNIIMPGLYVRVLMDOVAVDNAFVVPQOAV--TRGAKDTVMIVNAOG 335
Db      280 NAVYYIARSPFVNPDPGLATGKTONTVEIDGVKNVLIIPSLTYKRNKGRAFAVRLGADG 339
Oy      336 GMEPREVTVAOQOGTNNIVTSGLKDGDKRVVEGIGSIA 372
Db      340 KAVEREIRTKGMDSMNTEVKSGLKEGDKVVISETIAA 376

RESULT 13
AY38884
ID      AAY38884 standard; Protein; 392 AA.
XX
AC      AAY38884;
XX
DT      08-OCT-1999 (first entry)
XX
DE      Neisseria gonorrhoeae antigenic protein encoded by ORF85.
XX
KM      Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX      treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX
OS      Neisseria gonorrhoeae.
XX
PN      MO9924578-A2.
XX
PD      20-MAY-1999.
XX
PF      09-OCT-1998; 98WO-IB01665.
XX
PR      01-SEP-1998; 98GB-0019016.
PR      06-NOV-1997; 97GB-0023516.
PR      14-NOV-1997; 97GB-0024190.
PR      18-NOV-1997; 97GB-0024386.
PR      27-NOV-1997; 97GB-0025158.
PR      10-DEC-1997; 97GB-0026147.
PR      14-JAN-1998; 98GB-0000759.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR      MPI; 1999-327407/27.
XX      N-PSDB; AA212306.
XX
PT      Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX      diagnosis, treatment and prevention of infection
XX
PS      Claim 4; Page 430; 524pp; English.
XX
CC      Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX      and N. gonorrhoeae antigenic proteins. They are encoded by open
XX      reading frames (ORFs) AA21972-Z1358. The antigenic proteins,
XX      their fragments, their nucleic acids and antibodies are used for

```

RESULT 10

ABG18262
ID ABG18262 standard; Protein; 627 AA.
AC ABG18262;
XX
XX
XX
DT 18-FEB-2002 (first entry)
XX
XX
DE Novel human diagnostic protein #18253.
XX
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
XX
PN WO200175067-A2.
XX
XX
PD 11-OCT-2001.
XX
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS82449.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX
PS Claim 20; SEQ ID No 48621; 103bp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 627 AA;

Query Match 15.9%; Score 320.5; DB 22; Length 627;

Best Local Similarity 34.5%; Pred. No. 2e-19; Indels 19; Gaps 1;

Matches 81; Conservative 38; Mismatches 97;

QY 34 GGGAGRAPPAPVGVVTHPQVALTVELPGRLESTADVAQVGG110KRLFOBGSY 93

Db 385 GGRGMSGSLAPVQAATAVAEQAVPRVLTGTTTAANTVTVRSRQGLIALHFOBQO 444

QY 94 VRAQPLVQIDSSYTEANLESARQALATQAOTLAKADADLARYPLVAAEVVSQETDAA 153

RESULT 11

Db 445 VKAGDLAEIDPSOFKVAIAQAQGLAKDKATLANARRDLARYOQLAKTNLVSQELDAQ 504
QY 154 VTAKRSAPAGVKAQAQAIAKSGAGINLNSRITAPISGFGSKVSEGLTINAGDTTLAT 213
Db 505 QALVSETEGTITKADEASVASAQLOLDMSRITAPVDGRVGLQVGVNGINSSGDTTGIV 564
QY 214 RQTNPMYVNTVQASAEVWKL-----RQIABGLAADGVI 249
Db 565 TQTHPIDLVFTLPESDIATVVAQAKGKPLVEAMDRTNSKLTSEGLTISIDNQI 619

RESULT 11

ABP79725
ID ABP79725 standard; Protein; 392 AA.
XX
XX
AC ABP79725;
XX
XX
DT 07-MAR-2003 (first entry)
XX
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 5980.
XX
XX
KW Antibacterial; infection; vaccine; gene therapy.
XX
XX
OS Neisseria gonorrhoeae.
XX
XX
PN WO200279243-A2.
XX
XX
PD 10-OCT-2002.
XX
XX
PF 12-FEB-2002; 2002WO-IB02069.
XX
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
XX
PA (CHIR-) CHIRON SPA.
XX
XX
PI Fontana MR, Pizza M, Maignani V, Monaci E;
XX
XX
DR WPI; 2003-058415/05.
DR N-PSDB; AB240695.
XX
XX
PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection
XX
XX
PS Disclosure; Page 624; 815bp; English.
XX
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
XX
SQ Sequence 392 AA;

Query Match 15.5%; Score 313; DB 24; Length 392;

Best Local Similarity 28.0%; Pred. No. 4.7e-19; Indels 54; Gaps 10;

Matches 111; Conservative 63; Mismatches 169;

QY 7 KAMRAALAAVAVLVSSCGKGGAPAGGQAPAGREAPAPVGVVTHPQVALTVELPGR 66

Db 3 KMMKMAVAVAVA-----AVWGWGWSYKPEPQAAYITETVARGDISRTVSATGE 52

QY 67 LESIRKTDVRAQVGGITQKRLFOBGSYVRAQPLVQIDSSY-----EANLESARQ 119

Db 53 ISPSNIVSVAQASGOIKKLKYKLGQOVKKGGDIAEINSTTQNTIDMERSKLETVOAKL 112

QY 120 ATAQATLAKADADLARYPLVAAEVVSROEYDAVTAKRSAPAGVKAQAQAIAKSGINL 179

Db 113 VSAQIALGSAEKYKROALIMKDATSKEDLESAQDLAAAKANVLELKLINQSKISTIN 172

QY 180 R-----SRITAPISGFGSKVSEGLTINAGDTTLATIRQTNPMYVNTVQASAEVW 232

PN MO9927129-A1.
 XX
 PD 03-JUN-1999.
 XX
 XX 25-NOV-1998; 98WO-US25247.
 PF 25-NOV-1998; 97US-006517.
 XX
 PR 25-NOV-1997; 97US-006517.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 XX Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
 PI Rahme LG, Tan M, Tsongalis J;
 PI WPI, 1999-357851/30.
 XX
 DR Virulence factors useful in developing disease treatments
 XX
 PT
 PS Disclosure; Fig 14B; 228pp; English.
 XX
 XX The present sequence represents a *Pseudomonas aeruginosa* polypeptide
 CC sequence. *P. aeruginosa* is an opportunistic human pathogen present in
 CC soil water and plants. The specification describes virulence polypeptides
 CC and nucleic acid sequence encoding such polypeptides. These sequences
 CC can be used to identify a compound which is capable of decreasing the
 CC expression of a pathogenic virulence factor. Compounds that inhibit
 CC the expression or activity of virulence factor polypeptides can be
 CC used to treat pathogenic infections, especially where the infection
 CC is a *P. aeruginosa* infection.
 CC
 CC note: the sequences given in the specification were poorly legible, and
 CC in some instances assumptions were made as to the identity of the
 CC residue; it is therefore possible that the sequence given below is
 CC not entirely correct.
 XX
 XX Sequence 202 AA;
 XQ

[illegible]

PD 11-OCT-2001.
 XX
 PE 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS81885.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 20; SEQ ID No 48057; 103bp, English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probe,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/publ/published_pct_sequences.

SQ Sequence 703 AA;

Query Match	16.8%	Score 339	DB 22	Length 703
Best Local Similarity	28.7%	Pred. No. 5.6e-21		
Matches 87; Conservative	62	Mismatches 110	Indels 44	Gaps 5

QY 84 QKRLFQEGSYVRAGQPLYQIDSSTYEANLESARAQLATAQATLAKADA-----DLARY 136

Db 386 BEGLSÖYRAVDSRGLIIVADTQGHETREYENNAATYDSAKVICAPAAANIQLFTVNR 445
QY. 137 KPLVAEAESRÖEYDAVTAKRSAAEGVKQAQALIKAGILNRSRTAPISGFIOGSV 196

Db 446 QKLGTYISKQYDQALADAQOANAAVTAKAAVETAQINLLTPK--APLRLLAVALSNV 503

QY SGGTLLNAGTGTTVATIRPNNPMYVVNTOSASMMKLRRQIAEGKLIADGVIAVGIKPD 256
 : ::
 Db TEGALVONGATATATVQQDLPDIYDVDTQSNDPLRLKQELANGTLKQENCCKAKVSILTS 563

257 DGTVYPEKGRLLFADPVNNESTGQITLRAVPNDQNILMPGLYRVRLMDQVAVDNAFVVP 316

Db 564 DGKFPQDGTLEFSIVTVRQTGSTILRAIFPNPDHTLLPG-----604

QY 317 QQAVTRGAKDTVMIVNAQGMET---REVVAQQQSTNMWITVSGLD--GDKVVEGIS 370

Db 605 -----NIDWKNLAGFETFNHRIPLI SEIGTRWVLQTMQOHNTAGRLLPGSS 654

QY 371 IAG 373

Db 655 VSG 657

ABG15606
ID ABG15606 standard; Protein; 978 AA.
XX
AC ABG15606;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15597.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS79793.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 45965; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcf_sequences.
XX
SQ Sequence 978 AA;
XX
Query Match 22.8%; Score 459.5; DB 22; Length 978;
Best Local Similarity 30.6%; Pred. No. 2e-31;
Matches 132; Conservative 77; Mismatches 157; Indels 65; Gaps 12;
QY 11 AATAAATAATVLSGCGK--GGDMAAGGQAPAGREAPAP-----VGVVTVHP 54
DB 290 SGVIVKXIGLIMGCGQGRFGIDA--PVIREHQIDGRFQVGTGRHVVSQNVKXHC 345
QY 55 QTVALTVELPG-----RLBSLRTDVRAQVGGITQKRLFQSGSVVRAGQP---LYQID 104
DB 346 LAVNTRIKRTNIIAICRASQASCAVMGDDIDQPTARDHRSMTTARVSQSSVLABEL 405

QY 105 SSTYEANIESARAQATQAATLAK--ADADIA-----RYKPLVAAEAVSRQEV 150
DB 406 TRAVPCPEPLFRIPAAAGHPHAHCFDAELSLFWFNKDLIFRRFAKYVAA-----FW 459
QY 151 DAAVTAKGSAEAGYKAA-----QAATKSAGINLNKSRITAPISFGIOSKUSSEGLNAG 205
DB 460 RIASSSRSSASWRLRRAISPCTQPVVPTARNILAYTVATSPISGRIGSNVTEGALVONG 519
QY 206 DTTVLATIRQNPVNVVNTOSASSEVMKLRLQIAGCKLADGVALVAGIKPDGTVYPEKG 265
DB 520 QATLATVQQDLPYVDTOSNDPLRKQELANGTLKOENGAKAVSLITSDEKFPDGG 579
QY 266 RLTFADPVNVESTQITLRAAVPNDQNILMFGLYRVRLMDQAVADNAFVVPQAVTRGAK 325
DB 580 TLRFSDVTVDQGTSTILRAIFPNPDHTLLPGMFVRARLEEGNLPNAILVPOQGVTRPR 639
QY 326 D--TWMTVYNAAGMEPEEVYVAQCGGNNWYVTSGLKXGDKRVYEGISLA-----GITGA 377
DB 640 GDATVVLVVGADDKVETRPVVASQAIQDKWLTBGLKAGDRVVISGLQKVRPGVPGIWWA 699
QY 378 KKVT--PREVA 386
DB 700 EKLSELPASMS 710
XX
XX RESULT 6
XX AAB15924
XX ID AAB15924 standard; Protein; 464 AA.
XX
XX AAB15924;
XX
XX 05-OCT-2000 (first entry)
XX
XX E. coli proliferation associated protein sequence SEQ ID NO:281.
XX
XX Escherichia coli; E. coli; proliferation; inhibition; screening;
XX antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
XX Escherichia coli.
XX
XX WO200044906-A2.
XX
XX 03-AUG-2000.
XX
XX 27-JAN-2000; 2000WO-US02200.
XX
XX 27-JAN-1999; 99US-0117405.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Zykkind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2000-514822/46.
XX N-PSDB; AAA65929.
XX
XX Novel polynucleotides and polypeptides associated with microorganism
XX proliferation, used to identify inhibitors of bacterial growth and
XX proliferation, for use in antisense therapy -
XX
XX Claim 11; Page 205-206; 316pp; English.
XX
XX AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
XX sequences derived from Escherichia coli which inhibit E. coli
XX proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
XX nucleotide and protein sequences associated with E. coli proliferation.
XX AAA66056 and AAA66057 represent primers used for sequencing E. coli
XX proliferation inhibiting nucleotide inserts in an example from the
XX present invention. Methods from the present invention can be used to
XX identify a proliferation- required gene in a microorganism, by contacting
XX a microorganism with a proliferation- required gene activity inhibitory
XX nucleic acid identified in another organism, and determining if
XX inhibition occurs in the second microorganism. The nucleic acid sequences

PD 10-OCT-2002.
XX 12-FEB-2002; 2002WO-1B02069.
XX 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX MPI; 2003-058415/05.
DR N-PSDB; AB239062.
XX
XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a
PT medicament for treating or preventing *N. gonorrhoeae* infection -
XX
XX Disclosure; Page 386; 815pp; English.
XX
XX The present invention relates to proteins from *Neisseria gonorrhoeae*.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
XX
SQ Sequence 412 AA;

Query Match 96.2%; Score 1943; DB 24; Length 412;
Best Local Similarity 95.6%; Pred. No. 4.9e-163;
Matches 194; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAFYAFKMRRAAALAAVALVLSGCKGDAAGGAPGRAPAPVGVVTHPOTVALT 60
DB 1 MAFYAFKMRRAAALAAVALVLSGCKGDAAGGAPGRAPAPVGVVTHPOTVALT 60
QY 61 VELPGRLESLRTADVRAQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLA 120
DB 61 VELPGRLESLRTADVRAQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLA 120
QY 121 TQAQTLAAQADADLAAYKLVVAEAVSRQETPAANTAKSAEAGVYAAQAALIKSAGININR 180
DB 121 TQAQTLAAQADADLAAYKLVVAEAVSRQETPAANTAKSAEAGVYAAQAALIKSAGININR 180
QY 121 TQAQTLAAQADADLAAYKLVVAEAVSRQETPAANTAKSAEAGVYAAQAALIKSAGININR 180
DB 121 TQAQTLAAQADADLAAYKLVVAEAVSRQETPAANTAKSAEAGVYAAQAALIKSAGININR 180
QY 181 SRITAPISGFIQOSKVSSEGTLLNAGDTTVALTIRQTNPMYVNTOSASEVWKLRLROIAG 240
DB 181 SRITAPISGFIQOSKVSSEGTLLNAGDTTVALTIRQTNPMYVNTOSASEVWKLRLROIAG 240
QY 241 KLLAADGVIAVGIKEDGTIVPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYV 300
DB 241 KLLAADGVIAVGIKEDGTIVPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYV 300
QY 241 KLLAADGVIAVGIKEDGTIVPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYV 300
DB 241 KLLAADGVIAVGIKEDGTIVPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYV 300
QY 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYNAQSGMEPREVTVAAQOGTWMVITSGIKD 360
DB 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYNAQSGMEPREVTVAAQOGTWMVITSGIKD 360
QY 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYNAQSGMEPREVTVAAQOGTWMVITSGIKD 360
DB 301 RVLMDQVAVDVAFFVPOQAVTRGAKDTMTIYNAQSGMEPREVTVAAQOGTWMVITSGIKD 360
QY 361 GDKVVEGISTAGITGAKKVTPEKMASSENOAAAPQSGVOTASEKTKASEAE 412
DB 361 GDKVVEGISTAGITGAKKVTPEKMASSENOAAAPQSGVOTASEKTKASEAE 412

RESULT 4
ID AAY32854 standard; Protein; 397 AA.
XX AAY32854;
XX
XX 01-NOV-1999 (first entry)
XX
XX E. coli acra protein sequence.
XX
XX Acra; acrb; tolC; organic solvent resistance; indigo production;

KM steroid conversion; flooded crude oil treatment.
XX
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..24
FT Protein /note= "signal peptide"
FT /note= "mature acra"
XX
XX JF11221080-A.
XX
XX 17-AUG-1999.
XX
XX 09-FEB-1998; 98JP-0027537.
XX
XX 09-FEB-1998; 98JP-0027537.
XX
XX (MEIJ) MEIJI SEIKA KAISHA LTD.
XX
XX MPI; 1999-520718/44.
DR N-PSDB; AA211064.
XX
XX Escherichia coli having organic solvent resistance - useful for
PT production of indigo, conversion of steroid and treating flooded
PT crude oil
XX
XX
XX Claim 5; Page 8-9; 14pp; Japanese.
XX
XX This sequence represents the acra protein. The invention relates
CC to a method for the preparation of E. coli having organic solvent
CC resistance or of increased organic solvent resistance including a step of
CC transforming the E. coli with at least one gene of acra, acrb and tolC.
CC The E. coli can be used for the production of indigo from indol,
CC conversion of a steroid such as cholesterol and treatment of flooded
CC crude oil.
XX
XX
SQ Sequence 397 AA;

Query Match 37.8%; Score 762.5; DB 20; Length 397;
Best Local Similarity 42.2%; Pred. No. 9.3e-59;
Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;

QY 11 AALAAVAALVLSGCKGDAAGGAPGRAPAPVGVVTHPOTVALFVELPGRLESL 70
DB 11 AALAAVAALVLSGCKGDAAGGAPGRAPAPVGVVTHPOTVALFVELPGRLESL 70
QY 71 RTADVRAQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLAQATLAKAD 130
DB 71 RTADVRAQVGGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLAQATLAKAD 130
QY 64 RIAEVRPOVGGIILKRNFKESGSDIEAGVSLYQIDPATYQATYDSAKGDLAKAQAANVIAQ 123
DB 64 RIAEVRPOVGGIILKRNFKESGSDIEAGVSLYQIDPATYQATYDSAKGDLAKAQAANVIAQ 123
QY 131 ADLARYFVAAEAVSRQETPAANTAKSAEAGVYAAQAALIKSAGININR 190
DB 131 ADLARYFVAAEAVSRQETPAANTAKSAEAGVYAAQAALIKSAGININR 190
QY 124 LTVNRVYKLLGTQYISKQETDQALADQANAAVTAARAKAVERTAIRINLAATKYVSPISGR 183
DB 124 LTVNRVYKLLGTQYISKQETDQALADQANAAVTAARAKAVERTAIRINLAATKYVSPISGR 183
QY 191 IGQKVSSEGTLLNAGDTTVALTIRQTNPMYVNTOSASEVWKLRLROIAGLAAQVIA 250
DB 191 IGQKVSSEGTLLNAGDTTVALTIRQTNPMYVNTOSASEVWKLRLROIAGLAAQVIA 250
QY 184 IGKSNVTEGALVQNGQATLALATVQQLDPIYVDVTOSSNDPLRLKQELANGTLKENGAK 243
DB 184 IGKSNVTEGALVQNGQATLALATVQQLDPIYVDVTOSSNDPLRLKQELANGTLKENGAK 243
QY 251 VGIFEDDTIVPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYVRLMDQVAVD 310
DB 251 VGIFEDDTIVPEKGRLLFADPVVNESTGOITLRAAVNDONILMPGLYVRLMDQVAVD 310
QY 244 VSLITSDGIRFPQDGTLEFSDVTVDTGTSITLRAIPFNDHTLLPCMFARARLEGLNP 303
DB 244 VSLITSDGIRFPQDGTLEFSDVTVDTGTSITLRAIPFNDHTLLPCMFARARLEGLNP 303
QY 311 NAFVPOQAVTRGAKD-TMTIYNAQSGMEPREVTVAAQOGTWMVITSGIKDQKVVVEG 368
DB 311 NAFVPOQAVTRGAKD-TMTIYNAQSGMEPREVTVAAQOGTWMVITSGIKDQKVVVEG 368
QY 304 NATLVPOQGVTRTPRGDAVLVVGADDKVETRPVIAQIGDKVLTGKAGGRVVISG 363
DB 304 NATLVPOQGVTRTPRGDAVLVVGADDKVETRPVIAQIGDKVLTGKAGGRVVISG 363
QY 369 ISLAGITGAKKVTPEKMASSENOAAAPQSGV 400
DB 369 ISLAGITGAKKVTPEKMASSENOAAAPQSGV 400
QY 364 L-----QVNRPGVQVKAQEVTAADNNQDA--SGAQ 391
DB 364 L-----QVNRPGVQVKAQEVTAADNNQDA--SGAQ 391

RESULT 5

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:52:56 ; Search time 44 Seconds
(without alignments)
1486.258 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019
Sequence: 1 MAFYAFKXMRRAALAAVAL.....AAPGSGVOTASEAKTASAE 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_13Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2019	100.0	412	21	AAV90283
2	1989	98.5	412	23	AAU91064
3	1943	96.2	412	24	ABP78092
4	762.5	37.8	397	20	AAV32854
5	459.5	22.8	978	22	ABG15606
6	432	21.4	464	21	AAAB15924
7	432	21.4	464	22	AAAG96971
8	370.5	18.4	202	20	AAV29280
9	339	16.8	703	22	ABG17698

10	320.5	15.9	627	22	ABG18262
11	313	15.5	392	24	ABP79725
12	308	15.3	392	20	AAV38883
13	308	15.3	392	21	AAV38884
14	308	15.3	392	21	AAV74869
15	308	15.3	392	21	AAV74871
16	308	15.3	392	23	AAU72963
17	308	15.3	392	23	AAV73013
18	271	13.4	334	20	AAV38882
19	271	13.4	334	21	AAV74870
20	217	10.7	102	23	ABP31381
21	188	9.3	394	24	ABP79084
22	175	8.7	659	22	AAV3061
23	169	8.4	309	22	AAU34575
24	169	8.4	309	22	AAU38257
25	165.5	8.2	89	23	ABP04272
26	164.5	8.1	175	21	AAAB4578
27	164.5	8.1	175	23	ABP54530
28	161.5	8.0	285	22	AAU36320
29	158	7.8	115	23	ABP31959
30	158	7.8	262	23	AAAG80026
31	156.5	7.8	390	22	AAU35553
32	156.5	7.8	399	24	AAE30466
33	151.5	7.5	399	21	AAV81743
34	151.5	7.5	399	24	ABU01173
35	151	7.5	645	22	ABG21616
36	140	6.9	684	22	ABG17838
37	138.5	6.9	355	22	AAAG98387
38	137.5	6.8	444	20	AAV55920
39	133	6.6	1201	22	ABB61629
40	132.5	6.6	392	20	AAV38881
41	132	6.5	387	23	AAO17576
42	130	6.4	2453	21	AAAB12454
43	129.5	6.4	512	21	AAAG31336
44	129	6.4	4572	19	AAV52845
45	127.5	6.3	329	19	AAV71499

ALIGNMENTS

RESULT 1
AAV90283 standard; Protein: 412 AA.
AC AAV90283;
DT 24-OCT-2000 (first entry)
DE N. meningitidis BASB055 protein sequence.
KW BASB055; diagnosis; microbial infection; invasive bacterial disease;
KW Neisseria meningitidis infection; upper respiratory tract infection;
KW bacteraemia; meningitis; therapy.
XX Neisseria meningitidis.
OS WO2000043517-A1.
PN 27-JUL-2000.
PD 19-JAN-2000; 2000WO-EP00425.
PF 22-JAN-1999; 99GB-0001462.
PR 29-JAN-1999; 99GB-0002069.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA Thonard J;
XX WPI; 2000-476199/41.
DR N-PSDB; AAA37645.
XX

Novel human diagno
N. gonorrhoeae ami
Neisseria meningit
Neisseria gonorrhoe
Neisseria meningit
Neisseria meningit
Neisseria meningit
Neisseria meningit
Neisseria meningit
Human glycoprotein
N. gonorrhoeae ami
C. glutamicum prote
E. coli cellular p
Salmonella typhi c
Human ORF932 prote
Virulence gene pro
Actinobacillus ple
Pseudomonas aerugi
Human ORF932 prote
A. cryptum Alub pr
Haemophilus influe
H. influenzae emrA
Streptococcus pneu
S. pneumoniae type
Novel human diagno
Novel human diagno
Escherichia coli p
Pseudomonas fluore
Drosophila melanog
Neisseria meningit
M. catarrhalis MC41
HNRCR protein sequ
Arabidopsis thaliana
A. mediterranei ri
Helicobacter polyp

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X14199; CAA32413.1; -
 DR PIR; S02387; BVERCD.
 DR InterPro; IPR006143; HLYD.
 DR InterPro; IPR006144; HLYD_FAMILY.
 DR InterPro; IPR003997; RCD.
 DR Pfam; PF00529; HLYD.1.
 DR PRINTS; PR01490; RYTXOXIND.
 DR PROSITE; PS00543; HLYD_FAMILY.1.
 DR Hemolysis; Transport; Transmembrane; Inner membrane.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 75 POTENTIAL.
 FT DOMAIN 76 440 PERIPLASMIC (POTENTIAL).
 SO SEQUENCE 440 AA; 4801 MW; B786328A92DD0D48 CRC64;

Query Match 6.1%; Score 122.5; DB 1; Length 440;
 Best Local Similarity 24.1%; Pred.No.1.9; Mismatches 156; Indels 103; Gaps 20;
 Matches 100; Conservative 56;

QY 13 ALAAVALVYSSCGKGAAAGGAPAGREAPVGVVTVHPQVALTELPGRLSIRT 72
 DB 63 ALAAG-ALWWSVVGK-----TEIVHAAGKVPVVGSKIIAASETGRVAVLV 109
 QY 73 ADVRAOVGIIQKRLFOEGSYVRAGQPLVQIDSTYEANLESARAQLATAQATLAKADAD 132
 DB 110 AD-----NSRYAAGDVLLRLDAGVTEHEKRVQAAQAR-----QD 146
 QY 133 LARYKPLVAEAVSRQEYDAAVTAKRSAGVKAQAIAKS--AGINLRSRITAPIS-- 188
 DB 147 EARSRAMIRALDTGR---APVLAELPADPGMMAQSYLDSQYADYQALRSIEAATATY 202
 QY 189 -----GFTGQSVSEGTLLNAGDT-----TVLATIRQTNPMYVNTQSASQVM 231
 DB 203 RRDVGLVTLQIAHARGLRDDGVSQAAYLEKEQARMTEGRLROSE-----AQRALQ 256
 QY 232 KLRRQIAEGKLIADGVIAGVIGKFDGDTVYPEKGRLLFPADPV--VNESTGQITLRAAVPN 289
 DB 257 QTRQAFETLVLRK--LAAQAEQEIARTSAQRSLVLTAFVGVQQLVALTEGTAVAA 314
 QY 290 DONILM-----PGLYVRVLMQVAVDNAFVVPQAATRGADTVMI-----VNAQGMGEP 340
 DB 315 TQPLMNVVPSGAGIQVQQLD--SKDIGF-----VRAGAPATVKVGAVDYTKYGTLEGK 366
 QY 341 -----EYTVAAQOGTWTIVTSGK-----DG-DKVVVEGIST-AGI-TGAKKV 380
 DB 367 VLYVSPDIVDDROOHSYRTIALAHPALVVDGKPRLLKEGMVAQVADIRTSRRL 421

Search completed: September 8, 2003, 14:01:16
 Job time : 20 secs

AC Q52969; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical sensor-like histidine kinase R01002 (EC 2.7.3.-).
 GN R01002 OR SMC00059.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 NC NCBI_TaxID=382;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RC MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Gordin T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Fumelle B., Rampeger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.,
 RA "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 RN
 RP SEQUENCE OF 1-87 FROM N.A.
 RC STRAIN=RCR2011 / SU47;
 RC MEDLINE=95276304; PubMed=7756693;
 RA Keller M., Roxlau A., Weng W.M., Schmidt M., Quandt J., Niehaus K.,
 RA Jording D., Arnold W., Puehler A.;
 RT "Molecular analysis of the Rhizobium meliloti muck gene regulating the
 RT biosynthesis of the exopolysaccharides succinoglycan and
 RT galactoglycan.";
 RL MoJ. Plant Microbe Interact. 8:267-277(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Contains 1 histidine kinase domain.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL591785; CAC45574.1; -;
 DR EMBL; L37353; AAA74242.1; -;
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR004358; Bact_sens_pr_C.
 DR InterPro; IPR003661; His_kinase.
 DR InterPro; IPR005467; His_kinase.
 DR InterPro; IPR001610; PAC_kinase.
 DR InterPro; IPR000700; PAS-assoc_C.
 DR Pfam; PF02518; HATPase_C; 1.
 DR Pfam; PF00512; HisKA_1.
 DR PRINTS; PR00344; BCTRLENSOR.
 DR Pfam; PF00785; PAC; 1.
 DR SMART; SM00387; HATPase_C; 1.
 DR SMART; SM00388; HisKA_1.
 DR SMART; SM00386; HisKA_1.
 DR PROSITE; PS50109; HIS_KIN_1.
 DR PROSITE; PS50113; PAC; 1.
 KW Hypothetical protein; Sensory transduction; Transferase; Kinase;
 KW Transmembrane; Phosphorylation; Complete proteome.
 FT TRANSMEM 36
 FT TRANSMEM 56
 FT TRANSMEM 61
 FT TRANSMEM 173
 FT DOMAIN 246
 FT MOD RES 249
 FT MOD RES 249
 SO SEQUENCE 525 AA; 55899 MW; 5366357A463F99F0 CRC64;

QY 1 MAFYAFKAMRAA-----LAAVALVSSCGKGD-----AAGCGQPAQR 40
 DB 26 IARDFFRLRAVAAVSLTALPIVPLATLALPVSAALPAGAAALWASASLLAAAAAIAAGR 85
 QY 41 E-----APAP-----VGVTVHQTVALTYELGR-----LESRTADV 75
 DB 86 EPTDGEVEFAPAPLPLDNLNAYDLFAGLVTVH-DTRGHVLSVHGRDASEYLLKMRDPHG 144
 QY 76 RAQVGIIQKRLPQEGSVYRAGQPLYQIDSSSTYEANLESAPQAQATQAATLAKADADLAR 135
 DB 145 R---GFIEQIHVSRIAFIRA-----IDS---LRDSEK---SANDIRLEKTSADQPO 188
 QY 136 Y-----KPLVAE---AVSRQYDAAVTAKSAEAGVAAQAATKAGININRSRT 184
 DB 189 FAHYCEWTPLRDAEGNLLAIVAQSRDVEEARLQAEAAKAAHA--ESA--NDAKTRFL 244
 QY 185 A-----PISGTIGQSKVSEG-----TLNAGDTYVLAITRQ----- 215
 DB 245 AAVSHELRTPNLNAILGFSVDVLAGYFQKLENDROREYVSLIHQSGTHLLSVVNTMLDMSK 304
 QY 216 -----TTPMTVYNTQSASEVM-----KLRQIAEGKLAAD----- 246
 DB 305 IEAGRYELLEPRFYAEALACEMLSHQAREKGVRLTSRVSVE---INADQRAQO 361
 QY 247 ---GVIAVGIKFDDGTVPYPERGRLLFADPVVNESTGQITLRAAVPNQNIIMPGLYVRVL 303
 DB 362 ILINILGNALIKFTD-----RGLVTVDAALLEGMLKLTV-----SDPTGIIADDKLQML 410
 QY 304 MDQVAVDNAFVPPQAVTRGAKDT---VMYNAQSGHEPRVTV--AQOQGTNMIYVSG 358
 DB 411 -----GQPFVQIQNDYTRRYEGTGLSLVGLAEHLGGDVSIRSAEGESTVIVWITP- 463
 QY 359 KDQGVVVEGISIGITGAKVTPEKWAASSENOAAP-----OSGVQTASRAK 406
 DB 464 SDASGAARQCADAPVTEPPRLKQHDGKREAGVPSALHTGELIGREGHGGAQAK 523
 QY 407 TA 408
 DB 524 TA 525

RESULT 40
 CYAD BORPE STANDARD; PRT; 440 AA.
 ID CYAD BORPE.
 AC P11091;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cyad protein.
 GN CYAD.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=18323;
 RC MEDLINE=89091151; PubMed=2905265;
 RX Glaser P., Sakamoto H., Bellailou J., Ullmann A., Danchin A.;
 RA "Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-
 RT haemolysin bifunctional protein of Bordetella pertussis.";
 RL EMOO J. 7:3997-4004(1998).
 CC -1- FUNCTION: CYAD IS NECESSARY FOR TRANSPORT OF CALMODULIN-SENSITIVE
 CC ADENYLATE CYCLASE-HEMOLYSIN (CYCLOLYSIN).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U00039; AAB18563.1; --
 DR EMBL; AE000436; AAC76610.1; --
 DR PIR; S47807; S47807.
 DR EcGene; EG12290; y1aV.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD. 1.
 KM Hypothetical protein; Transmembrane, Signal; Complete proteome.
 FT SIGNAL 1 15
 FT CHAIN 16 378
 FT TRANSMEM 27 47
 FT TRANSMEM 57 77
 SO SEQUENCE 378 AA; 41771 MW; 66833C1E129BC3E51 CRC64;
 Query Match 6.2%; Score 125.5; DB 1; Length 378;
 Best Local Similarity 20.2%; Pred. No. 1.1;
 Matches 65; Conservative 49; Mismatches 112; Indels 95; Gaps 11;
 QY 53 HPQV---ALTVLPGRLSRLTADYRAVQVGGIIQKRLFGSGYVRAGQPIYQIDSSYIE 109
 DB 51 HPYFKAKQKAVISIP-----VVPQVGVVIEVTDKKNTLIKKEGVLFRLDPTRYQ 100
 QY 110 ANLESARAQLATAQ-----ATLAKADADLARYKPLVAEA----- 144
 DB 101 ARVRLMADIVTAIEKKKALGAELDEMAANTQAKATDKAKYKARYARSQAKVNPFS 160
 QY 145 ---VSROEYDAAVTAKRSAAEYKAKQAQAIRK-----AGININRS 181
 DB 161 ERDIDVARQNYLAQBAVSKSAAEQKIQSQDLSVLGEHSQIALKQAQLEAKYVNEQT 220
 QY 182 RITAPISGFIQSKYSEGTLLNAGDTTTLATIRQTNPMYVNVVQASAVMRLRQIAEKG 241
 DB 221 IVRAPSDGYVTQVIRPQT-----YASLPLRPVMPVIRIPQOKRIVAQFQNSILR 271
 QY 242 LLADGVTAIVGIRKPD--DGYVPEKGRLLFPADPVV---NSTGOITLRAAVP----- 288
 DB 272 LAPGDDAEV---FNALPGKVF--SGKLAISPAPGAYSTGTLQTLNTPASDDVIA 326
 QY 289 -----NDQNILMPGLYVRV 302
 DB 327 TIEDEHTDLSALPDGIYAQV 347
 RESULT 38
 APRE_PSEAE STANDARD; PRT; 432 AA.
 ID APRE_PSEAE STANDARD; PRT; 432 AA.
 AC 003025;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline protease secretion protein apRE.
 DE APRE OR PA1247.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OC NCB1_TaxID=287;
 RN NCB1_TaxID=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=93051361; PubMed=1427098;
 RA Duong P., Lazdunski A., Cami B., Murgier M.;
 RT "Sequence of a cluster of genes controlling synthesis and secretion
 RT of alkaline protease in Pseudomonas aeruginosa: relationships to
 RT other secretory pathways";
 RL Gene 121:47-54(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gabler R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody R.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -1- FUNCTION: INVOLVED IN THE SECRETION OF ALKALINE PROTEASE.
 CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X64558; CAA45856.1; --
 DR EMBL; AE004554; AAC04636.1; --
 DR PIR; G83489; G83489.
 DR PIR; S26697; S26697.
 DR InterPro; IPR006143; HLYD.
 DR InterPro; IPR006144; HLYD_FAMILY.
 DR InterPro; IPR003997; RtxD.
 DR Pfam; PF00529; HLYD. 1.
 DR PRINTS; PR01490; RTXTOXIND.
 DR PROSITE; PS00543; HLYD_FAMILY. 1.
 KM Transmembrane; Inner membrane; Transport; Complete proteome.
 FT DOMAIN 1 14
 FT TRANSMEM 15 36
 FT DOMFLCT 37 432
 FT CONFLCT 378 379
 SO SEQUENCE 432 AA; 48052 MW; 4C845C65DECF15EE CRC64;
 Query Match 6.1%; Score 123; DB 1; Length 432;
 Best Local Similarity 21.6%; Pred. No. 1.8;
 Matches 79; Conservative 56; Mismatches 99; Indels 132; Gaps 17;
 QY 57 VALTVLPGRLSRLTADYRAVQVGGIIQKRLFGSGYVRAGQPIYQIDSSYIEANLESAR 116
 DB 42 VPAIVTIIISGGRKS-----VOHPLGCVKHLTVRQDGEVGEPLIRMEPTQARANVDSL 96
 QY 117 AQLATQATLAKADADLARYKPLVAEAVSRQYDAVTKRSAE--AGYKAAQAARAKSA 174
 DB 97 NRYANARLNOARLQA-----EYD-----GRRTLEMPAGL-AEQAPLPTL 134
 QY 175 G--ININR-----SRITAPISGFIQ---SKVSEGT-----LIN-----A 204
 DB 135 GERLEIKORQLHSRQTLANELSLRANIEGLRAQLEGRQTBONORLOQRLNSQLSGA 194
 QY 205 GD-----TTVLATIRQTNPMYVNVVQASAVMRLRQIAEGLKLLA----- 245
 DB 195 RDLAEGSYMRNQLLEGQRQLAEVNAIRLSSSGRFGQIRQSIADQAQRIARREBEYKRV 254
 QY 246 DGYIA-----VGIK-PDQGVVPEKGRLLPAD 271
 DB 255 NGOLAETQVNAARTLWELLSARAYELRHAETRAPVSGYVAGIKYFTDDGVI----- 304
 QY 272 PVNNESTGOITLRAAVVNDQNIIMPGLYVRVLMDOV---AVDNAFVVPQAAVTRGAKDT 327
 DB 305 -----GPGEL-LWYIVNDSLEVEGQLAVNLVDRHSGLPVEMLTFAPNOSTKTPRVTGE 358
 QY 328 VMIYNA 333
 DB 359 VTMVSA 364
 RESULT 39
 YA02_RHIME STANDARD; PRT; 525 AA.
 ID YA02_RHIME STANDARD; PRT; 525 AA.

DR Pfam: PF00529; HlyD: 1.
 DR PRINTS: PRO1490; R1XTOKIND.
 KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil.
 FT TRANSMEM 9 28 POTENTIAL.
 FT DOMAIN 148 207 COILED COIL (POTENTIAL).
 FT NON TER 338 338
 SQ SEQUENCE 338 AA; 37548 MW; 5F56F234A9B42FEE CRC64;

Query Match 6.3%; Score 128; DB 1; Length 338;
 Best Local Similarity 23.0%; Pred. No. 0.68;
 Matches 73; Conservative 43; Mismatches 101; Indels 100; Gaps 13;

QY 60 TVEPLGRLESLRTADYRAQVGIIIOKRLFOEGSYRAQGPLYQIDSYEANLESARA-- 117
 DB 36 TTTTIGNIDIRQVQAAPDNGRLDLRV-OEGDRVKKQQLADLDPVRFQDAVPKDAVW 94
 QY 118 -----QLATAQATLAKADADLARYKPLVAEASRQEVAA 153
 DB 95 PRKRCWPAKRVVPRKSPRRGPKPAQAQTLNNAETTWQOQALARQYVPKQSLDA 154
 QY 154 VTAKRSAGVYAAQAAI-----KSAGINLAN-----SRITAPI 187
 DB 155 AALAKTARANDRAQQAALTLAIKGRKEDIAAARQQLADKAGLSLARRELDRLVAP 214
 QY 188 SGFTGQSVSEGTILNAGD---TTVLTITQTNPMYVNTQSASEVKKLRQIAEGKLL 243
 DB 215 DG-----VVODRILEPDMVSPQTPFTLALDNPVWV-----RAYLPEKAL- 255
 QY 244 AADGVIAVGIFKFDGTV---YPER---GRLLFADPVV--NESTGQIT-----LRA 285
 DB 256 ---GQVRLGKM---ATISSDSFGPKSPFGWVFISPTAEFTPKVTOTELTELVRV 309
 QY 286 AVPDONTLMPGLYVRV 302
 DB 310 YACNPQRLRLGMPVTV 326

RESULT 36
 AN36_HELPY STANDARD; PRT; 329 AA.
 ID AN36_HELPY
 AC P94851;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 36 kDa antigen.
 GN HPI488.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteriaceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Hp921023;
 RA Hocking D., Rothel L., Doidge C., Radcliff F., Lee A., Webb E.;
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RA MEDLINE=97394467; PubMed=9252185;
 RX Tomb J.-F., White O., Kellavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wattey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori."
 RL Nature 388:539-547(1997).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: AEO00647; AAO08523.1;
 DR EMBL: U86610; AAB81283.1;
 DR PIR: H64705; H64705.
 DR TIGR: HPI488;
 DR InterPro: IPR00397; RecD.
 DR PRINTS: PRO1490; R1XTOKIND.
 KW Antigen; Transmembrane; Complete proteome.
 FT TRANSMEM 11 31 POTENTIAL.
 FT VARIANT 67 67 H -> R (IN STRAIN HP921023).
 FT VARIANT 256 256 E -> D (IN STRAIN HP921023).
 FT VARIANT 272 272 K -> R (IN STRAIN HP921023).
 FT VARIANT 274 274 T -> A (IN STRAIN HP921023).
 SQ SEQUENCE 329 AA; 36199 MW; B33A7572705DF87 CRC64;

Query Match 6.3%; Score 127.5; DB 1; Length 329;
 Best Local Similarity 22.8%; Pred. No. 0.7; Indels 67; Gaps 6;
 Matches 63; Conservative 38; Mismatches 108;

QY 63 LPGRLESLRTADYRAQVGIIIOKRLFOEGSYRAQGPLYQIDSYEANLESARA----- 117
 DB 38 LQGFLEA-REYVSASKVGRIEKVFVKGDHKKGLDVFSSPELAKLAQAEAGKAA 96
 QY 118 -----QLATAQATLAKADADLARYKPLVAEASRQEVAA 153
 DB 97 KALDEYKGRSRETTINSARDWQAQKQATLAKETRYKQVDDYDNGVASLQKRDENVAA 156
 QY 136 YKPLVAEASRQEVDAV-----TAKRSAGVYAAQAAIKSAGINLANSRTAPISGF 190
 DB 157 YESTKYNESAYQKTKMALGASSESKIAKAKESAPALGOVNEVESYLVKQVATAPIDG 216
 QY 191 IGQKSVSEGTILNAGD---TTVLTITQTNPMYVNTQSASEVKKLRQIAEGKLL 240
 DB 217 VSNVTLISGELSPKGFVVMIDLKQSMWLKISPEKYNLFYKGER-BEYIYALKKSTK 275
 QY 241 ---KLAADGVIAVGIFKFDGTVPEKGRLLFADPV 273
 DB 276 FRVYKYLISVMDPFTWKATNNSNTYDMKSYEVEAIP 311

RESULT 37
 Y1AV_ECOLI STANDARD; PRT; 378 AA.
 ID Y1AV_ECOLI
 AC P37683;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein y1av precursor..
 GN Y1AV OR B3586.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes."
 RL Nucleic Acids Res. 22:2576-2586(1994).
 CC -I- SIMILARITY: STRONG, TO E.COLI Y1BH; SOME, TO P.AERUGINOSA APPE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

```

CC -1- SUBUNIT: Interacts with HDAC7. Forms a large corepressor complex
CC that contains SIN3A/B and histone deacetylases HDAC1 and HDAC2.
CC This complex associates with the thyroid (TR) and the retinoid
CC acid receptors (RAR) in the absence of ligand.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named Isoforms=2;
CC   Name=Long;
CC   IsoId=O60974-1; Sequence=Displayed;
CC   Name=Short;
CC   IsoId=O60974-2; Sequence=VSP_003411;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DOMAIN: The N-terminal region contains repression functions that
CC are divided into three independent repression domains (RD1, RD2
CC and RD3). The C-terminal region contains the nuclear receptor-
CC interacting domain that are divided in two separate interaction
CC domains (ID1 and ID2).
CC -1- DOMAIN: The two interaction domains (ID) contain a conserved
CC sequence referred to as the corn box. This motif is required and
CC sufficient to permit binding to unliganded TR and RARs. Sequences
CC flanking the CORN BOX determine nuclear hormone receptor
CC specificity.
CC -1- SIMILARITY: Contains 1 SANT-A domain.
CC -1- SIMILARITY: Contains 1 MYB-like domain.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U35312; AAB17125.1; -
CC EMBL; U22016; AAC52168.1; -
CC PIR; S60254; S60254.
CC TRANSFAC; T04688; -
CC MED; MGI:1349717; Ncor1.
CC InterPro; IPR001005; MYB_DNA_binding.
CC Pfam; PF00249; myb_DNA-binding; 2.
CC SMART; SM00717; SANT; 2.
CC PROSITE; PS50090; MYB 3; 1.
CC KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
CC Coiled coil; Alternative splicing.
CC FT DOMAIN 174 216 COILED COIL (POTENTIAL).
CC FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
CC FT DOMAIN 299 328 COILED COIL (POTENTIAL).
CC FT DNA_BIND 437 482 SANT-A (POTENTIAL).
CC FT DNA_BIND 501 550 MYB.
CC FT DOMAIN 606 616 COILED COIL (POTENTIAL).
CC FT DOMAIN 2073 2077 PRO-RICH.
CC FT DOMAIN 2277 2281 CORN BOX OF ID1.
CC FT DOMAIN 58 64 CORN BOX OF ID2.
CC FT DOMAIN 593 602 POLY-GIN.
CC FT DOMAIN 1044 1047 POLY-PRO.
CC FT DOMAIN 1713 1718 POLY-ALA.
CC FT DOMAIN 1968 1979 POLY-SER.
CC FT VASAPLIC 2333 Missing (in isoform Short).
CC FT CONFLICT 1952 1952 I -> T (IN REF. 2).
CC FT CONFLICT 2090 2090 A -> P (IN REF. 2).
CC SQ SEQUENCE 2453 AA; 270640 MW; 52208B40382F7E6A CRC64;
CC -----
CC Query Match 6.4%; Score 130; DB 1; Length 2453;
CC Best Local Similarity 21.3%; Pred. No. 4.9;
CC Matches 106; Conservative 70; Mismatches 164; Indels 158; Gaps 26;
CC -----
CC QY 28 GDAAGC-----GDPAGREAPRVGVVYVYVQVATLVLPGRLSLRTAD---- 74
CC Db 1066 GGSISGCTPGTYLSHQAQYQOEAPKPSVG-----SISGLPRQESTKAAPLTYI 1116

```

```

CC QY 75 -----VRAQVGIIQ--KRLFOEGSVYRAGQPLYQIDSTYEANLES 114
CC Db 1117 KOEEFSPRQNSQPEGILVAAGHGGVVRGTAAGVQESIR--GTPAKISVET-----ISS 1171
CC QY 115 ARAQL-----ATQA--TLAKADADLARYKPL-----VAEAVSR----- 147
CC Db 1172 LRSGITQGTALPQACIPTEALVKGPVSR--MPIEESSEPKYREBAASKHYIEGKSGHI 1230
CC QY 148 QEYDAVATAR-----SAEAGVKAQAARKSAGINLRSRITAPISGT----- 191
CC Db 1231 LSYDNIKNAREGTRSPPTAHEMSLKSYEAVEGSIHQ--GMSMBESPVASPLEGLICPALP 1289
CC QY 192 ---GQSRYSEGLTLNAGDTIVLATIRQTNMYVYVVTQSAEVMKLRQIAEGK--LHAAD 246
CC Db 1290 RGSFPHSLKERTYLS-----GSIMGTRP--ATAESFPDGUKYRQIIRRESPPIAFAE 1340
CC QY 247 GTIAYGIRKFDGTVYPERKRLP-----ADPVNVEST---GQITLRAPV 288
CC Db 1341 GAITKGKPYDGIITTIKEMGRSHEIPRODILTQESRKTPEVOSTRPIEGSISQGTPIK 1400
CC QY 289 NDQN-----ILMPGLYVRVLMDOVAVDNAFVVPQ--QATRGAKOTVMVNA 333
CC Db 1401 FDNNSGQSALKNVKSLLTGSPSKLPROMLE-----IYVENIKYVERKAYED--VKA 1449
CC QY 334 OGGMPEPEVTVAAQOQGNMIVTGLKDGDKVVEGISIAGI--TGAKKVPKWEASSENQ 391
CC Db 1450 GEFYRARRHSVSSGSR--VLRSTLHEAPKAQIS---PELYDSSARRPIFVQYNTISR 1503
CC QY 392 AAAPQSGVGTASEAKTAS 409
CC Db 1504 GSPMNRRTSDVSSSKSAS 1521
CC -----
CC RESULT 35
CC YBHG THIFE STANDARD; PRT; 338 AA.
CC ID YBHG THIFE
CC AC Q9L9D4;
CC DT 26-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Hypothetical protein in asrc 5' region (ORF1) (Fragment).
CC OS Thibacillus ferrooxidans.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
CC OC Acidithiobacillaceae; Acidithiobacillus.
CC OX NCBI_TaxID=920;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=ATCC 33020;
CC RX MEDLINE=20250620; Pubmed=10788346;
CC RA Butcher B.G., Deane S.M., Rawlings D.E.;
CC RT "The chromosomal arsenic resistance genes of Thibacillus ferrooxidans
CC have an unusual arrangement and confer increased arsenic and antimony
CC resistance to Escherichia coli."
CC RL Appl. Environ. Microbiol. 66:1826-1833(2000).
CC CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (Potential).
CC CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC CC -1- CAUTION: Comparisons with its orthologs suggests that this protein
CC may be full-length.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF173880; AAF69246.1; -
CC DR HMAAF; NF_01304; - 1 -
CC DR InterPro; IPR006143; H1XD.
CC DR InterPro; IPR003997; RtxD.

```

OY 378 KKATPKWASSENQAAAFQS 397
 Db 514 LLITAEAMADIPIRGSPQS 533

RESULT 33

EMRK_ECOLI STANDARD; PRT; 387 AA.
 ID EMRK_ECOLI P76517; P76944; P76945;
 AC P52599; P76517; P76944; P76945;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 26-FEB-2003 (Rel. 41, Last annotation update)
 DE Multidrug resistance protein K.
 GN EMRK OR B2368.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Uesumi R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alta H., Baba T., Hayaashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Sato Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamaguchi S., Horiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, D78168; BA11236.1; ALT INIT.
 CC EMBL, AE000325; AAC75427.1; -
 CC EMBL, D90867; BA16239.1; ALT INIT.
 CC EMBL, D90866; BA16231.1; -
 CC PIR, E65010; E65010.
 CC EcoGene, EG13282; emrk.
 CC InterPro, IPR005694; Emr.
 CC InterPro, IPR006143; HLYD.
 CC InterPro, IPR006144; HLYD_FAMILY.
 CC Pfam, PF00529; HLYD; 1.
 CC TIGRFAMs, TIGR00998; Ba0101; 1.
 CC PROSITE, PS00543; HLYD_FAMILY; FALSE_NEG.
 CC Transport; Transmembrane; Inner membrane; Complete proteome.
 CC DOMAIN 1
 CC TRANSMEM 16 CYTOPLASMIC (POTENTIAL).
 CC FT 17 POTENTIAL.

FT DOMAIN 38 387 PERIPLASMIC (POTENTIAL).
 SQ SEQUENCE 387 AA; 42585 MW; ABAD26F5D68FF38F CRC64;
 Query Match 6.6%; Score 133.5; DB 1; Length 387;
 Best Local Similarity 19.4%; Pred. No. 0.37;
 Matches 61; Conservative 53; Mismatches 106; Indels 95; Gaps 9;

OY 75 VRAVGGIIOKRLFOEGSVYRAQOPLYQIDSTYEANLESAR----- 116
 Db 57 ISAVSGSVTVVNHKDTNRYQGDILVSLDKTATIALNKAKNLNIIVRTNKLVLQDK 116
 OY 117 ---AQLTAQATLAKADADLAARKPLVAEAVRQEV---DAVTKRAEAGVKAQA 169
 Db 117 QYSAEVASARIQYQSLDEYNNRVPYLAQGVISKETLEHTKDTLISKALNAIQYKA 176
 OY 170 -----AIKSAGININRRIPTAFISPTIGSQSVSEGLTNAGDTT 208
 Db 177 NKALWMTPLNRPQVVEADATKEMWALKRTDISPVGYIAQNSVQGVYSPQS- 235
 OY 209 VLATIRQTNPVYVNTQASAEVKKLRQIAEGKLAAAGVIAVGIRPDGTVPKGRRL 268
 Db 236 -LMAVVPARQMWNANPKETQLTDVR--IGSVNIISD-----LYGE----- 274
 OY 269 FADPVVNESTGQTLTAAVNDONILMPGLYRVYVLDQVAVDNAF-VVPOQAVTRGAKDT 327
 Db 275 -----NVVPHG---RVTGINMGTAFAFSLPQNAATGMWIKI 308
 OY 328 VMTVNAQGMREPV 342
 Db 309 VQRPVEVSLDPKEL 323

RESULT 34

NCRL_MOUSE STANDARD; PRT; 2453 AA.
 ID NCRL_MOUSE
 AC Q60974; Q60812; 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor co-repressor 1 (N-COR1) (N-COR) (Retinoid X receptor
 DE interacting protein 13) (RIP13).
 GN NCOR1 OR RXRIP13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RC TISSUE=Pituitary;
 RX MEDLINE=96008539; PubMed=756614;
 RA Hoerlein A.J., Naeser A.M., Heinzel T., Torchia J., Gloss B.,
 RA Kurokawa R., Ryan A., Kamei Y., Soederstrom M., Glass C.K.,
 RA Rosenfeld M.G.;
 RT "Ligand-independent repression by the thyroid hormone receptor
 RT mediated by a nuclear receptor co-repressor.";
 RL Nature 377:397-404(1995).
 RN [2]
 RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Liver;
 RX MEDLINE=95280959; PubMed=7760852;
 RA Seol W., Choi H.S., Moore D.D.;
 RT "Isolation of proteins that interact specifically with the retinoid X
 RT receptor: two novel orphan receptors.";
 RL Mol. Endocrinol. 9:72-85(1995).
 RN [3]
 RP INTERACTION WITH HDAC7.
 RX PubMed=10984530;
 RA Downes M., Ordentlich P., Kao H.-Y., Alvarez J.G.A., Evans R.M.;
 RT "Identification of a nuclear domain with deacetylase activity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:10330-10335(2000).
 CC -1- FUNCTION: Mediates the transcriptional repression activity of some
 CC nuclear receptors by promoting chromatin condensation, thus
 CC preventing access of the basal transcription.


```

RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoevel F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puchler A.,
RT "The complete sequence of the 1,603-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL603645; CAC49273.1; -.
DR PIR; A95951; A95951.
DR HAMAP; MF_01304; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
DR Plasmid; Complete proteome.
FT TRANSMEM 5 23 POTENTIAL.
FT DOMAIN 101 208 COILED COIL (POTENTIAL).
FT DOMAIN 184 193 POLY-ALA.
SQ SEQUENCE 334 AA; 35680 MW; 9B5D1617B5E78FC3 CRC64;

Query March 6.8%; Score 136.5; DB 1; Length 334;
Best Local Similarity 22.3%; Pred. No. 0.21;
Matches 71; Conservative 46; Mismatches 109; Indels 93; Gaps 10;

QY 62 ELPLRLSLR-----TADVRA-----OVGIIQRLPFGSGSYRAGPLVQIDST 107
DB 22 DLPLRLGPRPRQAVLYGVNDIRQVSLGFRVSGRIAEIARVDEGDSVAGVIAKLDMP 81
QY 108 YEANIESRAQALATRAQATLAK-----ADADLA--RYKPLVA 141
DB 82 FROAVGAEEAEAEVYRATLTLKLRAGARAAETIAQARATHEERLAELENKALFEERAKOLRP 141
QY 142 AEAASROEYDAAVTAKRSAGVKAQAAL-----KSAG 175
DB 142 NGTISQAEILDQANASRAADAPARSAREALVLEEGNPAEDIAAAAAQANATKADSAR 201
QY 176 INLNRSRTABISGFIQSGKYSE--GTLINAGDTTVALIRQTNPMYVNVQASQVWKLK 234
DB 202 ISLSDTQLAARSDGYI--LSRVREIGALVSPAD--IVVYLSLLEPVW-----R 246
QY 235 RQIABGKLAAAGVIAVGIRKFDGTVPEKGRLLPADVYNVESTGQI-----TL 283
DB 247 AYVABPDLGVRVHPKMKVTITSDTAPDRIVEGVGFIISVAEFTPKSVTPELARTDLYRL 306
QY 284 RAAVNDQNIIMPGLYVRV 302
DB 307 RIVIANPKGLRQGMPTV 325

RESULT 32
CH63_RHIME
ID CH63_RHIME STANDARD; PRT; 544 AA.
AC Q93070.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60-kDa chaperonin 3 (Protein Cpn60 3) (groEL protein 3).
GN GR0L3 OR GR0L3 OR RA0064 OR SMA0124.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

```

```

OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federici N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymb megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007200; AAK64722.1; *.
DR PIR; H95269; H95269.
DR HAMAP; MF_00600; -.
DR InterPro; IPR001844; Chaperonin Cpn60.
DR InterPro; IPR002423; Cpn60/TCE-1.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONIN3 Cpn60; 1.
DR Chaperone; ATP-binding; Heat shock; Multigene family; Plasmid;
DR Complete proteome.
SQ SEQUENCE 544 AA; 57521 MW; 2590163503E4479A CRC64;

Query March 6.7%; Score 134.5; DB 1; Length 544;
Best Local Similarity 21.4%; Pred. No. 0.48;
Matches 94; Conservative 69; Mismatches 190; Indels 87; Gaps 16;

QY 7 KAMFAAALAAVALVSSCGKGDAAQGGPAGREAPVGVTVHPQVALTVLPLGR 66
DB 132 KARATKVISSEIAOVGTIANAGDAGVEMIRAMEKVGNEGVTV----- 177
QY 67 LESRTADVAAQVGGIIQKRLPFGSGSYRAGPLVQIDSTYENNESAA-----RAQLAT 121
DB 178 -EEARTADTELDV--VEGQDFRG-YL--SPFVTNAERMRVELDDPYLLIHEKRLGS 229
QY 122 AQAATLAKADLARYKPLVAEAASVROEYDAAVTAKRSAGVKAQAALIKSAGINLRS 181
DB 230 LQANLPLTEAAVQSGKRLIITSEDEGEVALTVVNR-LNGGLKI--AAVTPFGGRK 286
QY 182 RITAPISGFIQSGKYSE-----TLNAGDTTVL-----ATIR-QT 216
DB 287 AMLEDIAVLTAGQMSIEDLGIKLENTVLLDMIGRARVLIKDTTITIDSGDKKSIGARV 346
QY 217 NPMYVNVQASQVWKLKRLQIABGKLAAAGVIAVGIRKFDGTVPEKGRLLPADPVNE 276
DB 347 SQIQAQIEETASDDEKQLQERLAKLAGVAIVRWG-GATELEVKEKKDRI--DDALNA 402
QY 277 STGQITLRAAVPNDQNIIMPGLYVRVLMDOAV-----DNAFV-----VPQ 317
DB 403 T-----RAAV-----ESGIVGGVALLRAKSAIVGLTDNADVAGISIVRALEAPIRQ 453
QY 318 QAVTRGAKDTVMIVNAAGQKMEPREVTVAAQGGTNWIVTSGLKDGDKVYVBCISAGITGA 377
DB 454 IADNAGVGSIVVQKLVKDGHDHNGQFDAQETVYDMIKAGIVPAKVVRALRDAGSIAS 513

```



```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alpa H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Ogihara T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113 (1997).
RN [4]
RP REVIEW.
RX MEDLINE=94262163; PubMed=8203018;
RA Lewis K.;
RT "Multidrug resistance pumps in bacteria: variations on a theme."
RL Trends Biochem. Sci. 19:119-123 (1994).
CC -1- FUNCTION: THE EMR LOCUS CONFERS RESISTANCE TO SUBSTANCES OF HIGH
CC HYDROPHOBICITY. EMRA PROBABLY PARTICIPATE IN A TRANSPORT SYSTEM
CC TO EXTRUDE TOXINS AND DRUGS FROM THE CELL.
CC -1- SUBCELLULAR LOCATION: Inner membrane-bound.
CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
CC
CC * This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86657; AAA3724.1; -.
DR EMBL: AE000353; AAC75732.1; -.
DR EMBL: D90891; BAA16547.1; -.
DR EMBL: D90892; BAA16552.1; -.
DR PIR: F65048; F65048.
DR EcoGene: EG11354; emrA.
DR InterPro: IPR005694; Emr.
DR InterPro: IPR006143; HLYD.
DR InterPro: IPR006144; HLYD_FAMILY.
DR Pfam: PF00529; HLYD_1.
DR TIGRfams: TIGR00998; 8a0101; 1.
DR PROSITE: PS00543; HLYD_FAMILY; PLASE_NEG.
KW Antibiotic resistance; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 46 POTENTIAL.
FT DOMAIN 47 390 PERIPLASMIC (POTENTIAL).
FT CONFLICT 62 62 I -> M (IN REF. 1).
FT CONFLICT 138 138 K -> O (IN REF. 1).
SQ SEQUENCE 390 AA; 42736 MW; 0FB9A86C0A270F9 CRC64;

```

Query Match 7.5%; Score 150.5; DB 1; Length 390;
 Best Local Similarity 21.4%; Pred. No. 0.038;
 Matches 67; Conservative 53; Mismatches 108; Indels 85; Gaps 11;

```

OY 75 VRAVGGLIQRLQOBSYVAGOPLYOISSTYEANIESRAQLATN-----122
DB 64 IMSVGSVTVKVMADNTDFVEKGVLTLDPTDARQAFKAKTALASSVROTQMLNSK 123

```

```

OY 123 -----OATLAKADADLARYKELVAEAVNSROEYDAVTAKSAA-----162
DB 124 OLQANIEVQKIALAKASDYNRRVPLGNANLIGSEELQHRDATSQAQDVAIQYNA 183
OY 163 -----GVKAQDAIKSAGININRSRTAPISGFIGOSKVSSEGLINAGDTT 208
DB 184 NQAMITLQKLEDDPAPVOQAATEVAMWALALERRITLSPMTGYVSRVAVPGCAQISP-TTP 242
OY 209 VLAITRTQNMVYVNVVQSASEVWKLRRQIAEGKLLAD-----GVIAVGKIPDGT 259
DB 243 LMAVPAVIN-WWVDANFKEIOIANMR--ICQPVTTITTDYGDVYKTKV-VGLDMGTGS 298
OY 260 VPEKGRLEPADPVNVESTQITLRAAVPNDONILMGLVYRVLMQDVAVDNAFVVPQA 319
DB 299 AF-----SLT---PAQNAATGWIKVQRLP-----VRIELDKQL-----EQY 333
OY 320 VTREKADTWTIVN 332
DB 334 PLRIGSLTVSVN 346

```

RESULT 29

```

YIBH_ECOLI STANDARD; PRT; 378 AA.
ID YIBH_ECOLI
AC P32107;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUN-1998 (Rel. 36, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Hypothetical protein yibH.
GN YIBH OR B3597 OR Z5021 OR EC54473.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93259920; PubMed=8387990.
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
RT "Rns elements of Escherichia coli K-12: complex composites of shared
RT and unique components that have different evolutionary histories."
RL J. Bacteriol. 175:2799-2808 (1993).
RN [2]
RP REVISION TO 205.
RC STRAIN=K12;
RA Hill C.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H., Burland V., Daniels D.L., Plunkett G. III, Blatner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586 (1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grodeck E.O., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

```



```

Db      161 LKSAQDKLSQYRTGNREODIAQAASLEQAQAQADLQDTLLAPANGTLITRAVE 220
Qy      198 EGTLLMAGDTVLTATIRQTNEMV-----NTVQSASEWKLKRLQIAEGTLAADGVIA 250
Db      221 PGMNMGSTVL--TLSTLRVWRAVYVDERNLSOT-----QPGRDILLTYTDG--- 266
Qy      251 VGIRFDGTVYPER---GRLLFADPVVNESTGQIT-----LRAAVPDQ 291
Db      267 -----RPDKRYHKGIFVSP-----TAEFTPKVTPDRLDLYVRLRIIVDAD 311
Qy      292 NLMPLGLYRVLMDOVA 308
Db      312 DALRQGMPTVTKFNDDEA 328

RESULT 25
CUSB_ECOLI
ID CUSB_ECOLI STANDARD; PRT; 407 AA.
AC P77239;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative copper efflux system protein cusb precursor.
GN CUSB OR B0574.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federapfel N., Hyman R., Kaiman S., Komp C., Kuri O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RT DNA Res. 3:137-155 (1996).
RN [4]
RP GENE NAME.
RC STRAIN=K12 / DH5-alpha;
RX MEDLINE=20461235; PubMed=11004187;
RA Munson G.P., Lam D.L., Outten F.W., O'Halloran T.V.;
RT "Identification of a copper-responsive two-component system on the
RT chromosome of Escherichia coli K-12.";
RL J. Bacteriol. 182:5864-5871 (2000).
CC -1. FUNCTION. MAY BE A COMPONENT OF A COPPER EFFLUX SYSTEM.
CC -1. SIMILARITY. TO S. TYPHIMURITUM SLB.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL; AB000162; AAC73675.1; -
DR EMBL; U82598; AAB40772.1; -
DR EMBL; D90699; BAA35208.1; -
DR EMBL; D90700; BAA35214.1; -
DR PIR; D64790; D64790.
DR Ecogene; EG14235; cusb.
KM Transport; Copper; Signal; Complete proteome.
FT SIGNAL 1 POTENTIAL.
FT CHAIN 29 407 PUTATIVE COPPER EFFLUX SYSTEM PROTEIN
FT CUSB.
SQ SEQUENCE 407 AA; 44304 MW; E949390000C61FC6 CRC64;

Query Match 8.0%; Score 162; DB 1; Length 407;
Best Local Similarity 24.1%; Pred. No. 0.0083;
Matches 84; Conservative 51; Mismatches 141; Indels 72; Gaps 13;

Qy 50 VTVPQVALTVLPGRLSLRTADVRAQVGGIQRKL-FOEGSYRAGQPLVQIDSTY 108
Db 99 VTGPIPLFPAQSFPANVSYNEYQAIIVQARAAGFIDKVPPLTVGDKVKGTPLDLITPDW 158
Qy 109 -EANTESAPRAQLATAQTLAKA-----DADLARYKPLVAEAVSROEYDAVTA 156
Db 159 VEAQSEVLLRLREGTATQTEGILERLRLAGMPEADIR--LIATQKIQ----- 205
Qy 157 KRSAGAVKAAQAIAKAGINLNSRTTAPISGFIGSKVSEGLNAGDTTVLATIRQT 216
Db 206 -----TRFTLKAPIDGVTITAFDLRAG--NMIAKDNNVAKIQGM 241
Qy 217 NPMVYN--VTQSASEWKLKRLQIAEGKLLAADGVIAVGIFDDGTVYPERKRLFPADPV 274
Db 242 DPWVNTAIFESTAMLVKDSQF-----TLTPPAR-PDKTILIRKMTLL---PGV 287
Qy 275 NESTGQITLRAVVPNDNIMPLGLYRVLMDOVAVDNAFVVPQA-VTRGAKDTMTVNA 333
Db 288 DAATRTQLRLVEDNDADLKPQNMALQLN-TASEPMLLIPSGALIDTSGEQGVITVDA 346
Qy 334 QGGMEPREVTVAQQGNTNIVTSGLDGDKVYVEGI---SIAGITGA 377
Db 347 DGRFVPRVAVFQASQGVTAIRSGLAEGEKVSSGFLIDSEANISGA 394

RESULT 26
EMRA_HAEMIN
ID EMRA_HAEMIN STANDARD; PRT; 390 AA.
AC P44928;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Multidrug resistance protein A homolog.
GN EMRA OR H10898.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goodyear J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uettersack T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical membrane protein ybhg.
GN YBHG OR STY0853 OR T2070.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
EX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagers K.,
RA Krogh A., Latteen T.S., Leach S., Moute S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL627268; CAD05265.1; -
DR EMBL: AE016841; AAC69690.1; -
DR HAMAP: MF_01304; -; 1.
DR InterPro: IPR006143; HLYD.
DR Pfam: PF00529; HLYD; 1.
KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
KW Complete proteome.
RN TRANSMEM 5
RN DOMAIN 140 209 POTENTIAL.
RN SEQUENCE 331 AA; 36339 MW; 6D6583979B5F48 CRC64;
SQ
Query Match 8.3%; Score 168.5; DB 1; Length 331;
Best Local Similarity 22.7%; Pred. No. 0.0027;
Matches 72; Conservative 49; Mismatches 89; Indels 107; Gaps 10;
QY 70 LRTADVAVQVGGIIQKRLFOEGSYVRAGQPLYOIDSSTYEANLESAPAOATL--- 126
DB 41 IRTVNISFRVGGRLASLNVDGDTIKAGVYLGELDHAPYENALMQAKGVSVVAQYDLM 100
QY 127 -----AKADADLA-----RYKPLVAEAVERQEDAVATKRSADG 163
DB 101 LAGYRDEEIAQAAAVRAQAAVYVQNFYNRQOGLMKSKRTISANDLENASSRDOAQT 160
QY 164 VKAAQ-----AAIKSAGINLNSRITAPISGFGOSKVS 197
DB 161 LKSKQDRLISQYRTGNREODIAQAKASLEQAKAQIAQAOGLDQDTTLTAPAGTLLTAVE 220
QY 198 EGTLLNAGDTTATIRQTNPMV-----NTVQASAEVWKLRQIAEGTLAADVIA 250

```

```

DB 221 PGMNLMNGSTVL---TLSTRPVWRAVYDERNLSQT-----QPGRDILLYTDG--- 266
QY 251 VGKIFDQGVYPER---GRLLFADPVVNESIGQIT-----LRAAVPDQ 291
DB 267 -----RPDPYHKGICFVSP-----TAETPRTVETPDLRTLVRLRIITVDAD 311
QY 292 NIMPGLYVRVLMDOVA 308
DB 312 DALRGWPFVTKFNDDEA 328
RESULT 24
YBHG_SALTY STANDARD; PRT; 331 AA.
AC 082QFP0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical membrane protein ybhg.
GN YBHG OR STY0818.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT Lt2.";
RL Nature 413:852-856(2001).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE008733; AAL19755.1; -
DR StvGene: SG27272; ybhg.
DR HAMAP: MF_01304; -; 1.
DR InterPro: IPR006143; HLYD.
DR Pfam: PF00529; HLYD; 1.
KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
KW Complete proteome.
RN TRANSMEM 5
RN DOMAIN 140 209 POTENTIAL.
RN SEQUENCE 331 AA; 36310 MW; 6754949E5033B83 CRC64;
SQ
Query Match 8.3%; Score 168.5; DB 1; Length 331;
Best Local Similarity 22.7%; Pred. No. 0.0027;
Matches 72; Conservative 49; Mismatches 89; Indels 107; Gaps 10;
QY 70 LRTADVAVQVGGIIQKRLFOEGSYVRAGQPLYOIDSSTYEANLESAPAOATL--- 126
DB 41 IRTVNISFRVGGRLASLNVDGDTIKAGVYLGELDHAPYENALMQAKGVSVVAQYDLM 100
QY 127 -----AKADADLA-----RYKPLVAEAVERQEDAVATKRSADG 163
DB 101 LAGYRDEEIAQAAAVRAQAAVYVQNFYNRQOGLMKSKRTISANDLENASSRDOAQT 160
QY 164 VKAAQ-----AAIKSAGINLNSRITAPISGFGOSKVS 197

```

RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Samped G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000181; AAC73882.1; -
 DR EMBL; D90717; BAA35461.1; -
 DR EMBL; D90716; BAA35455.1; -
 DR PIR; C64816; C64816.
 DR Ecogene; EG13315; ybhG.
 DR HAMAP; MF_01304; -; 1.
 DR InterPro; IPR006143; H1YD.
 DR Pfam; PF00529; H1YD; 1.
 KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
 KW Complete proteome.
 FT TRANSMEM 7 26 POTENTIAL.
 FT DOMAIN 107 210 COILED COIL (POTENTIAL).
 FT SEQUENCE 332 AA; 36416 MW; A61DC1D96F84785E CRC64;
 SQ
 Query Match 8.5%; Score 171.5; DB 1; Length 332;
 Best Local Similarity 23.6%; Pred. No. 0.0018;
 Matches 71; Conservative 42; Mismatches 101; Indels 87; Gaps 7;
 QY 70 LRTADVRAQVGIIQKLFQEGSYVRAQGYQIDSTYEANLESARAQATL--- 126
 Db 42 IRTNLSRVRGGRVSLAVDGDALIKGQVGLDHRKVEIALMQAKGVSAQAQVDLM 101
 QY 127 -----AKDADIA---RYKPLVAEAVSROEYDAVTAQRSAG 163
 Db 102 LAGYRNEEIAQAAAVKQQAAYDAQNFYRQOGLMWSRTISANDLENASSRDOAQT 161
 QY 164 VKA-----AAAIKASGINRSRITAPISGFGOSVVS 197
 Db 162 LKSAQDKLRQYRSGNREDDIAQASLEQAQAQALNLDSTLAPSGLTLTRAVE 221
 QY 198 EGTLLNAGDTVLATIRQTNPMYVNTQASAEVVKLRQIAEGKLLAADGVIAVGKFD 257
 Db 222 PGTYLNEGCT--VFTVSLTRPVW-----RAYVDERNLDQAQPKRVLLVTDG 267
 QY 258 GTVYPEKGRLLFADPVNVESTGQIT-----LRAVPNDQIIMPLGLYR 301
 Db 268 RPDPRYHQIGFVSP---TAEFPKTVETPDRTDLVLRRLRVTDADALRQGMPEVT 322
 QY 302 V 302
 Db 323 V 323
 RESULT 22
 YDHT ECOLI STANDARD; PRT; 285 AA.
 AC P76185; P76893;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ydhT.
 GN YDHT OR B1644.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.

OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1245-1247(1997).
 CC
 CC SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kawai H., Kashimoto K., Kimura S., Kitakawa M., Kikugawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
 RA Samped G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000259; AAC74716.1; ALT_INIT.
 DR EMBL; D90807; BAA15404.1; -
 DR Ecogene; EG13944; ydhU.
 DR InterPro; IPR006143; H1YD.
 DR Pfam; PF00529; H1YD; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26 POTENTIAL.
 FT CONFLICT 284 285 GQ -> VNNRIYVWLAQFALVQCHAGAMALCTQYCHVS
 FT
 FT SEQUENCE 285 AA; 31496 MW; FC3032D13764A869 CRC64;
 SQ
 Query Match 8.4%; Score 169; DB 1; Length 285;
 Best Local Similarity 25.4%; Pred. No. 0.0021;
 Matches 71; Conservative 43; Mismatches 105; Indels 60; Gaps 9;
 QY 52 VHPQVALTVELPGRLESLTADVRAQVGIIQKLFQEGSYVRAQGYQIDSTYEAN 111
 Db 40 IBAQVSLTPQVSGRIVELNIKD-----NQLVNAAGDLLLTIDKTPQIA 83
 QY 112 LESARAQATRAQATLAAADADLARYKPLVAEAVSROEYDAVTAQRSAGVQAQDAI 171
 Db 84 ELNAQAQALAAQOSPLAANNEANRRRL--SQNTISAEELDTANLVNAPMAQASVDAQATL 142
 QY 172 KSAGINLNRSRTAPISGFGOSKVSGLTLNAGDTVLATIRQTNPMYVNTQASAEV 231
 Db 143 KQAQWOLAQTEIRAPVSGWVNTLTRIGDYADTG-----KPLFALVDSHSRYVI 191
 QY 232 -----KLRAQIAEGK---LAAAG-----YIAGIKRQDGTVPYKGRLLFADPVN 275
 Db 192 GYFEETKL-RHIREGAPAQTLIVSDNKTLOGHVSIGRAIYDQSVESDSSLI---PDVK 246
 QY 276 ESTGQITLRAAVPRNDQIIMPLGLYVRVLMQAVAVDNAFV 314
 Db 247 PNVPWVLAQRPV-----YRPALDKVPDVTLLV 274
 RESULT 23
 YBHG SALTI STANDARD; PRT; 331 AA.
 AC O82879;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ybhG.
 GN YBHG OR B1644.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.

	Matches	72,	Conservative	38;	Mismatches	101;	Indels	93;	Gaps	7
QY	70	LRTADVRAOVGGII	IOKRLFOEGSVYRAGQFL	OIDSSTYEANLESAPQAQLATQAOTLAKA	129	:	:	:	:	:
		:	:	:	:	:	:	:	:	:
Db	42.	IRTVLSLRSPVGGRVESLAVDEGDALIKAQVLGELDHKKPEYLALMQAKGVSAQ---	AQY	98						
QY	130	DADLARYPELVAAEAV-----	SROFYDAA	153						
		:	:	:	:	:	:	:	:	:
Db	99	DLMLAGYDEDEIATAQAAAQAAVAQAAAYDAQNPFYNQOGLMKSRTTISANDLENAFNRSSRDQA	158							
QY	154	VTAKRSAE-----	AGVKAAQAIAKSAGINILNRSRTTAPISGFIGQS	194						
		:	:	:	:	:	:	:	:	:
Db	159	QATLKSAQDKLRQYRSGNRBODIAQAKASLEBAQAOQLAAQAEINLDOSTIVABSDGTILTR	218							
QY	195	KVSEGGTLLINAGDDTVLATIRTQTNPMYYVNVTOSASEVMKLROIARIEGKLLAADGVAIVAGIK	254							
		:	:	:	:	:	:	:	:	:
Db	219	AVEPCTVLNEGQT-VFTVSLTRLRPVV-----	RAVVDERNLDQAQPKRKVLLY	264						
QY	255	FDDGTIVYPEKGRLLFADRPVNAESTGOIT-----	LRAAVPDONTILMPGL	298						
		:	:	:	:	:	:	:	:	:
Db	265	TIDGRPNKRKYHQIGVSP-----TAIEFTPKTVERPEDLTDLVYRLRIVTTADADALROQM	319							
QY	299	YVRV	302							
		:	:							
Db	320	PVTV	323							

```
CC CC RESULT 20
AC YBHG_ECO57 STANDARD; PRT; 332 AA.
ID _YBHG_ECO57
AC O8Xf79;.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical membrane protein ybhgc.
GN YBHGC OR Z1015 OR EGS0873.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CX Enterobacteriaceae; Escherichia.
NX NCBI_Taxid=83334;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=1126551;
RA Pena N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kitchpatrick H.A.,
RA Postal G., Heckert J., Klinsk S., Boutin A., Shao X., Miller L.,
RA Groetbeck E.U., Davis N.W., Lim A., Djmalanta E.T., Potamoustas K.,
RA Aposteca J., Anantharaman T.S., Lin Y., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RN Nature 409:529-533(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Ikida T., Takami H., Honda T., Saekawa C., Ogatawara N., Yasunaga T.,
RA Kubura S., Shibata T., Hattori W., Shinagawa H.,
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12." ;
RL DNA Res. 8:11-22(2001).
--1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
(Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
```

```
CC or send an email to license@lsb.sib.ch).
CC -----
DR EMBL, AEO055260, AAG55166.1; -.
DR EMBL, AP002553, BAB34296.1; -.
DR PIR, A99738, A99738.
DR HAMAP, MF_01304; -.
DR InterPro, IPR006143, HLYD.
DR Pfam, PF00552, HLYD.1.
DR Hypothetical protein, Transmembrane, Inner membrane, Coiled coil;
KW Complete proteome.
KW TRANSMEM 7 26 POTENTIAL.
FT DOMAIN 108 210 COILED COIL (POTENTIAL).
SQ SEQUENCE 332 AA; 36417 MW; 75D1AB11C3089050 CRC64;
```

	Query Match	8.5%; Score 172.5; DB 1; Length 332;
	Best Local Similarity	23.7%; Pred. No. 0.0016;
	Matches	72; Conservative 36; Mismatches 101; Indels 93; Gaps 7;
QY	70 LRTADVAVQGGIIQKELFQEGSVVRAGPELYOIDSSTVEANLESAPALATQAATLAKA	129
Db	42 IRTVNLSFRVGGRVESLAVIDEDGAIKAGVYLGEIDHKPFYEIALMQAKAGVSVAQ---	AQY 98
QY	130 DADLARKEPLVAABAV-----	-SRQEIDAA 153
Db	99 DLMLAGRDEEILQAAAAVKAQAAVDYAQNFPYNRQGLWKSRTISANDLENASSRDAQ	158
QY	154 VTKARSAE-----	AGVKAAQAIKSKAGINLRNSRTITAPISGFIQS 194
Db	159 QATLKASQDKLRQYRGSGNREODIAQAKASILEQAOQLAQLELNLDQSTLLTAFSPDGTLLTR	218
QY	195 KVESEGTLLNAGDTTVLATIRQTNPVMYNVTOSASEWKLREROIAEGKLTAADVIAVGIK	254
Db	219 AVERGTYLANBGR--VFVTSLTRPVVV-----	RAYDERLDDAQPERKTLVLY 264
QY	255 FDDGYIVPEKGRIIFPADPVVNESGTQIT-----	LRAVPNDQNIIMPGI 298
Db	265 TDRPRDXPHYGQIGFVSP-----	TAEFTPKTEVTPDLRTDLVYRLRIIVTDADDAALRQM 319
QY	299 YWRV 302	
Db	320 PVTV 323	

RESULT 21			
ID	YBHG_ECOLI	STANDARD;	PRT; 332 AA.
AC	P75777;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Hypothetical membrane protein ybhG.		
GN	YBHG OR B0795.		
OS	<i>Escherichia coli</i> .		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; <i>Escherichia</i> .		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MG1655;		
RX	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,		
RA	Riley M., Colliado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of <i>Escherichia coli</i> K-12.";		
RL	Science 277:1453-1474(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12;		
RX	MEDLINE=97061202; PubMed=8905232;		
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,		
RA	Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kasumoto K.,		
RA	Kimura S., Kikugawa M., Makino K., Maeda S., Maki T., Mizobuchi K.,		


```

OY 154 VTAKSAEAGVKA-----OAAIKSAGINIRSRITAPI 187
D 152 RTANQARANLQAQKDLAQFLSGNRPQELAQAEANLAQTAEALAQQLNADITLLAPS 211
OY 188 SGFIGQSKVSEGLILNAGDTTLATITROTNPVYNVTOSASEVWKLRRQIAEGKLLAADG 247
D 212 AGTLTTRAVEPGLTILASANT--VFTVSLTDPVWV-----RAYVSERHQQAIP 257
OY 248 VIANIGFEDDGTVPPEK---GRLLPADPVNVESTGQIT-----LRANVP 288
D 258 GSEVEV-FTDGR--PDKEYHGKIGVSP---TAEFTPTKEVPDLRTDLVRLRIIT 309
OY 289 NDONILMPGLYVRVLMDO 306
D 310 DADESLRGMPVTVTRFPQ 327

RESULT 18
YJCR_ECOLI STANDARD; PRT; 343 AA.
AC P32716;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yjcr.
GN yjcr OR B4082.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blatter F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417(1993).
CC -1- SUBCELLULAR LOCATION: Inner membrane-bound (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00006; AAC43176.1; -.
DR EMBL; AE000481; AAD13465.1; -.
DR PIR; A65217; A65217.
DR EcoGene; EG11954; yjcr.
DR InterPro; IPR005694; Emt.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR TrEMBL; TIGR00998; 8a0101.1.
KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW Complete proteome.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 33 POTENTIAL.
FT DOMAIN 34 343 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 343 AA; 36908 MW; C599B870C475F5FE CRC64;

Query Match 8.8%; Score 178.5; DB 1; Length 343;
Best Local Similarity 23.9%; Pred. No. 0.00073;
Matches 87; Conservative 46; Mismatches 116; Indels 115; Gaps 12;

OY 6 FKAMRAALAAVAVLVSSCGKGGDAAGGQAPAGREAPVGVYVTHPQVATLVELPG 65
D 13 FPAALLVAL-ALVALVF-----VIVKVDSPSTNDAYASADITIDVPEVSG 57

```

```

OY 66 RLESRTADYRAQGVGIIQKRLPQESGYRAGOLYQIDSEYANLESARAQLA----- 120
D 58 RIVELATVDNQ-----VKQGLLFPIDRPYEPANLAKAASLAALDKQ 101
OY 121 -----TAQTLAKADA-----DLARYPELVAAEVSROEYDAATYAK 157
D 102 IMLTORSVDAQORGAOSVNATVEKARAAKQATDTLRTPELKEGFVSAEDVDRAATQ 161
OY 158 RSAEAGVKA-----AOAAIKSAG-----INLNSRITAPISFGIQ 193
D 162 RAADPADNAAVLLOQSHASAVSGVDALVAQRAVEADIALTKHLEMATVRAPPDGVIS 221
OY 194 SKVSEG-----TLNAGDTTLATITROTNPVYV-NVTOSASEVWKLRRQIAEGL 242
D 222 LKTSVGFASAMRPIFLIDTRHMYVIANFREDDLNKIRSGTATIRMSDSGTFEGKV 281
OY 243 LAADGVAVGIRKDDGTVPPEKRLFPADVNVESGQITLRAVPPDONILMPGLYVRV 302
D 282 -----DSIGYGLPDDG-----GLVIGLPEKVSRSINWVVAQRF-----VKI 320
OY 303 LMDQ 306
D 321 MVDK 324

RESULT 19
YBHG_ECOLI6 STANDARD; PRT; 332 AA.
AC Q8FJN6;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical membrane protein ybhg.
GN ybhg OR C0878.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Iacon S.-R., Boutin A., Heckert J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mody H.L.T., Domeneberg M.S., Blatter P.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF016757; AAN79351.1; -.
DR HAMAP; MF_01304; -.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
KW Hypothetical protein; Transmembrane; Inner membrane; Coiled coil;
KW Complete proteome.
FT TRANSMEM 7 26 Potential.
FT DOMAIN 108 211 Coiled coil (Potential).
SQ SEQUENCE 332 AA; 36402 MW; 75DBF761CEB98855A CRC64;

Query Match 8.6%; Score 173.5; DB 1; Length 332;
Best Local Similarity 23.7%; Pred. No. 0.0014;

```

```

DT 28-FEB-2003 (Rel. 41, last sequence update)
DE 28-FEB-2003 (Rel. 41, last annotation update)
DE Modulation protein nolF.
OS NOLF OR RA0481 OR SMA0876.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OS Plasmid pSymba (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
CN NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AK631;
RX MEDLINE=91360053; PubMed=1909418;
RA Baev N., Endre G., Petrovics G., Banfalvi Z., Kondorosi A.;
RT "Six nodulation genes of nod box locus 4 in Rhizobium meliloti are
RT involved in nodulation signal production: nodm codes for
RT D-glucosamine synthetase."
RL Mol. Gen. Genet. 228:113-124 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Hutzar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymba megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
CC -1- FUNCTION: INVOLVED IN THE PRODUCTION OF MEDICAGO-SPECIFIC
CC NODULATION SIGNAL MOLECULE.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FIVE
CC FRAMESHIFTS IN POSITIONS 78, 83, 119, 308 AND 329.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58632; CAA41486.1; ALT FRME.
DR EMBL; AE007238; AAK65139.1; -.
DR PIR; A95322; A95322.
DR InterPro; IPR006143; Hlyd.
DR Pfam; PF00529; Hlyd. 1
KM Plasmid, Nodulation; Complete proteome.
FT CONFLICT 9 L -> V (IN REF. 1).
FT CONFLICT 43 K -> E (IN REF. 1).
FT CONFLICT 82 E -> D (IN REF. 1).
FT CONFLICT 160 OL -> HV (IN REF. 1).
FT CONFLICT 189 A -> R (IN REF. 1).
FT CONFLICT 237 T -> M (IN REF. 1).
FT CONFLICT 301 A -> T (IN REF. 1).
FT CONFLICT 358 V -> A (IN REF. 1).
FT CONFLICT 363 II -> Y (IN REF. 1).
SQ SEQUENCE 367 AA; 39031 MW; 67793C426F21E9BF CRC64;

Query Match 9.1%; Score 184; DB 1; Length 367;
Best Local Similarity 23.0%; Pred. No. 0.00037;
Matches 88; Conservative 62; Mismatches 182; Indels 50; Gaps 12;

QY 11 AALAAVALVLSGCGKSGDAAGGPGAREAPAVGVTVHQTALVYELPGRLESL 70
DB 13 AFAITLAVTTVLSSQ-GRAIQVRRGSP-----IELAKADVSTAVRQDMANERVIVGSLTPI 67
QY 71 RTADVRAQVGGIITQKRLFOEGSVYRAGOPLYOIDSSTYEANLESARQLATAQTATLAKAD 130
DB 68 RRTITRSVSTIIEFLPVQIGDVNAGDLVFRGALLESANVGRKRAEADALSAQTLEAE 127
QY 131 ADIARXKPLVAEAESVROEYDAVTAKRSAEAGVKAQAIAKSGAINLSRITAPISGF 190

```

```

DB 128 AVLEERNTLREGASASATRLAALADVLRLAQRSLKQAEVSDERSLSHAENVAERFQGV 187
QY 191 IGQSVSEGLTLNAGDTTVALTIRQTPMRYNVQSQSEWKLK-RQIAEGLKLAADGVI 249
DB 188 IAAVSVEGQVPL-NTQMLTVELNRLVEDAGVPTSRPLTLKQSV-----LTVEG-- 240
QY 250 AVGIKFDGIVYPEKGLLPADPVNVESTQITLRAAVPNDQILMPGLVYRVLMDOVAV 309
DB 241 -----PGRTFSGEVARI-----SPTADAGSRARVFIADVNEGLLRGGMFT---IGDLRV 289
QY 310 D---NAVVYQAVTRAKQDTWIVNAGGMEREYVVAQOQGNW-----ITYSGKDG 361
DB 290 DDQKDVIALPASIRHDADGFVLKYRAGVLRPRVGL-----GRSMRDRLVOVSGVSEG 345
QY 362 D-----KVVVEGI 369
DB 346 DVIYTAPLPLVNVNTPIIIGI 367

RESULT 17
Y873 YERPE
ID Y873 YERPE STANDARD; PRT; 328 AA.
AC Q93AT7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Hypothetical membrane protein Ecs0873.
GN Ecs0873.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Yokohama / Biovar Antiqua;
RA Radnedge L., Agron P.G., Worsham P.L., Andersen G.L.;
RT "Genome plasticity in Yersinia pestis."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Membrane-associated. Inner membrane
CC (Potential).
CC
CC -----
CC -1- SIMILARITY: BELONGS TO THE UPF0194 FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF426171; AML27384.1; -.
DR HAMAP; MF_01304; -; 1
DR InterPro; IPR006143; Hlyd.
DR Pfam; PF00529; Hlyd. 1.
KM Hypothetical protein; Transmembrane; Inner membrane; Coiled coil.
FT TRANSMEM 7
FT DOMAIN 80 112 POTENTIAL.
FT DOMAIN 139 212 COILED COIL (POTENTIAL).
SQ SEQUENCE 328 AA; 35950 MW; C9B905A0C4ID36A9 CRC64;

Query Match 9.1%; Score 183; DB 1; Length 328;
Best Local Similarity 24.8%; Pred. No. 0.00038;
Matches 79; Conservative 48; Mismatches 97; Indels 94; Gaps 11;

QY 60 TWELPRLSLRTADVRAQVGGIITQKRLFOEGSVYRAGOPLYOIDSSTY-----EANL 112
DB 33 SLTLGVND-IRTVNIGFRVAGRLAIVDEGDIDHPGQTLGKLDGPYLNALKQAQNAV 91
QY 113 ESARADL-----ATAQTATLAKADADLARKPLVAEAESVROEYDA 153
DB 92 QSAQAQLALLKAGYREEETIAOVSEVAQROAAFDVADNPLKQOQGLMASKAVASANELENA 151

```

OS Alcaligenes eutrophus (Ralstonia eutropha).
 OG Plasmid pMOL28.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 CC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH34;
 RX MEDLINE=91139046; PubMed=8380802;
 RA Liesegang H., Lemke K., Siddiqui R.A., Schlegel H.-G.;
 RT "Characterization of the inducible nickel and cobalt resistance
 determinant chr from pMOL28 of Alkaligenes eutrophus CH34.";
 RL J. Bacteriol. 175:767-778(1993).
 CC -1- FUNCTION: NICKEL AND COBALT RESISTANCE PROTEINS CNRA, CNRB, CNRC
 CC CNRH AND CNRR MAY BE INVOLVED IN THE REGULATION OF CNR.
 CC -1- SIMILARITY: SOME, TO A. EUTROPHUS C2CB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M91650; AAA21969.1; -
 DR PIR; F47056; F47056.
 DR InterPro; IPR006143; H1YD.
 DR InterPro; IPR005695; Membrane_fus2.
 DR InterPro; IPR006311; Tat.
 DR Pfam; PF00529; H1YD; 1.
 DR TIGRFAMs; TIGR00999; 8a0102; 1.
 DR TIGRFAMs; TIGR01409; Tat signal seq; 1.
 KW Plasmid; Transport; Nickel; Cobalt;
 SQ SEQUENCE 395 AA; 40082 MW; 075C2165B5B8A09F CRC64;
 Query Match 9.6%; Score 194; DB 1; Length 395;
 Best Local Similarity 25.5%; Pred. No. 0.0001;
 Matches 100; Conservative 55; Mismatches 167; Indels 70; Gaps 14;
 QY 14 LAAVALVSSCGKGD--AAAGGAPAGREAP-----APV----- 46
 DB 13 MIAGVAAVAAVGAHLPVSEKSPASTQAPBAQKPOSAPFPGKLEKIPATYLAAN 72
 QY 47 -----VGVTVHPQTALVTELPGRLSELRDVRAGVGIIQKRLFOGSGSYRA 96
 DB 73 IAVEPVASAAVGTETLAPATA--ALPG-----SEAVTSRAAGAVQVRRLGDDVKA 124
 QY 97 GQPLEYQIDSTYEANLESARAQLATAQTLAKADADLARVYKPLVAEAESVROEYDAVTA 156
 DB 125 GDVLAALVDSF--EAGMAAEKRVNAQAKDLARKTYE--REASLFOQGVTPROEMAAKAA 180
 QY 157 KRSAAAGVKAQAQAIKSAGI-NLNRS-RITAPISGFIGOSKYVSEGLTNAGDTVLATIR 214
 DB 181 LDVAQAELRAATYAQAASHLASDGRSVAVSPFAGKITAQSVTLGAFV--APQAELFRVA 238
 QY 215 QTNMYVAVVTSQASAEVMTLRQIAEGKLADGVIAVGIRKPDGTVPERKLLFADPV 274
 DB 239 GTGAVQVEAAVTAADTSRIAG-SEATLLANG-----SFLSARVAQVPTV 284
 QY 275 NES--TGQITLRAAVPNDONILMPGLYVYVLMDOVAVDNAFVFPQAVTGAQKDTMIVN 332
 DB 285 TGSARVATVVVPAQPTDRLVYEGSVQR-LRTAVADAALASVBDANQNDGADVLFVR 343
 QY 333 AAGGMEPREVTVAAQOQGTNMTVTSGLKGDGVY 364
 DB 344 TQEGFRPMPVLYVTRSGGSAQILSGVQAGEOV 375

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Putative membrane fusion protein silB precursor.
 GN silB.
 OS Salmonella typhimurium.
 OG Plasmid pMG101.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=99128056; PubMed=9930866;
 RA Gupta A., Matsui K., Lo J.-F., Silver S.;
 RT "Molecular basis for resistance to silver cations in Salmonella.";
 RL Nat. Med. 5:183-188(1999).
 CC -1- FUNCTION: COMPONENT OF THE SIL CATION-EFFLUX SYSTEM THAT CONFERS
 CC RESISTANCE TO SILVER. MAY BE PART OF A THREE-COMPONENT
 CC CATION/PROTON ANTI-PORTER.
 CC -1- SIMILARITY: TO E. COLI CUSB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF067954; AAD11748.1; -
 DR KJ Transport; Antiport; Signal; Plasmid.
 FT SIGNAL 1 28
 FT CHAIN 29 430 PUTATIVE MEMBRANE FUSION PROTEIN SILB.
 SQ SEQUENCE 430 AA; 47719 MW; A3269C6D5F7F80F5 CRC64;
 Query Match 9.3%; Score 187; DB 1; Length 430;
 Best Local Similarity 23.3%; Pred. No. 0.0003;
 Matches 97; Conservative 70; Mismatches 168; Indels 82; Gaps 17;
 QY 19 ALVSSCGKGG-----DAAGGAPAGREAPAVGVTVHPQTALVTELPRLE 68
 DB 10 AIISSLIAGLISVTAMQYLNSSQKTVPBAQKPEKV-LFWYDPKMPDKPKGKSP 68
 QY 69 SL-----RTADYRA--QVGGI-----IQKRLFOGSGSYRAGQPLQIDSTYEANLES 114
 DB 69 FMDMDLVPKYADESGKSGGIRDPQVQ-NLGKTKQKXTRGLANT--SQITPAVSY 124
 QY 115 ARAQLATAQTLAKADADLARVYKPLVAEAESVROEYDAVTAKSAAE----- 162
 DB 125 NEVQFVIVQ--ARSDFEYKVPPLTIGDHVKGTPLIDITIPBWVAQSEFLLSGTG 181
 QY 163 ---GVKAQAALISAGI--NLNRS-----ITAPISGFIGOSKYVSEGLTNAGD 206
 DB 182 TPTQIKGVLERLRLRAGPREDIORLRSTRITQRTFKABIDGVTTFDIDRTG--NMISK 239
 QY 207 TVTIAITRQTNPMVNV--VTSQASEVWKLRLQIAEGKLADGVIAVGIRKPDGTVPER 264
 DB 240 DKVVAQIQGMDPWAISAAPESIAVLIKDSQPE-----ISVP-AYPDITFIVEK 288
 QY 265 GRLLFADPVNVESTQITLRAAVPNDONILMPGLYVYVLMDOVAVDNAFVFPQAVTGA 324
 DB 289 WNLV---PSVDQTRTLQVRLQVTKNDEFLKPGNNAVLKLNTQSQEMLLIPSGAVIDTGK 345
 QY 325 KQIVMYVNAAGGMEPREVTVAAQOQGTNMTVTSGLKGDGVYBEI---SIAGTGA 377
 DB 346 EQRVITVDEGKFPVKQIHVLHESQOOSGIGSLNEDGVVVSGLFLIDSEANITGA 402

RESULT 16
 NOLF_RHIME
 ID NOLF_RHIME STANDARD; PRT; 367 AA.
 AC P25196;
 DT 01-MAY-1992 (Rel. 22, Created)

RA Kunio T., Kusano T., Oyatsu H., Senoo K., Kanazawa S., Matsumoto S.;
 RT "Cloning and sequence analysis of czc genes in Alcaligenes sp. strain
 RL C14.";
 CC Biosci. Biotechnol. Biochem. 60:699-704(1996).
 CC -1- FUNCTION: CZCA AND CZCB TOGETHER WOULD ACT IN ZN(2+) EFFLUX
 CC NEARLY AS EFFECTIVELY AS THE COMPLETE CAC EFFLUX SYSTEM (CZCABC).
 CC THE CACB PROTEIN IS THOUGHT TO PUMP ZN(2+) CATIONS TO THE CZCA
 CC TRANSPORT PROTEIN.
 CC -1- SIMILARITY: SOME, TO A. EUTROPHUS CNRB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D67044; BA01060.1; -
 CC PIR: J04699; J04699.
 CC InterPro: IPR006143; HLYD.
 CC InterPro: IPR005695; Membrane_fus2.
 CC Pfam: PF00529; HLYD: 1.
 CC TIGRfams: TIGR00999; 8a0102; 1.
 CC Plasmid; Transport; Zinc; Cobalt; Cadmium resistance.
 CC KW SEQUENCE 520 AA; 54574 MW; AF7000ABA955C224 CRC64;
 SO
 Query Match 11.7%; Score 236; DB 1; Length 520;
 Best Local Similarity 25.7%; Pred. No. 4.8e-07;
 Matches 102; Conservative 57; Mismatches 166; Indels 70; Gaps 12;
 QY 9 MRALAAVALVLTSSCGKGDAAAGQPA-----GREAPAV----- 46
 Db 126 VRATGESQALKFVVS-----GDALSSQOPVAEPHFVDTANTVTLPGSSSPLAVRLSKEG 180
 QY 47 -----VGVV--TVHPQYVALTYELPG--RLSELRADPRAOVGIIQKRLPFG 91
 Db 181 KIELTADQALKTGVVQVQAGSAGKAVQAGVQFGEIRFNEDEKTAHVAVPRLAGVESVPANIG 240
 QY 92 SYVRAGOPLYQIDSSYEANLESARQATLTAQATLAKADADLARYKPLVAEAVSROEYD 151
 Db 241 QOVKKQGLVLAIVAST-----GLSDQSELLAAOKRLDLARVYDEKLEKMEKISADEYTL 296
 QY 152 AAVTAKSAEAGVKAQAQAIKISAGIN-----LNSRITAPISGFIGOSKVSSEGLTLNAGD 206
 Db 297 SARNAIDQAQISVQNAQOKLTALGASNSSTALNRYELAPAPGMIVKHIISLGEAV--AD 354
 QY 207 TTVLATIRQTPMVTNTQSAEYWKLRQIAEGKL--AADGYIANGIKRDDGTIVPEK 264
 Db 355 NANFTLSDLSVWAEFVSAKADVERVR--IGERKASINSASSDVYA-----DGTV----- 402
 QY 265 GRLEFADPVNVESTGQITLRAAVPNDONIILMPGLVYVRILMDQVAVDNAFVVPQQAIVRGA 324
 Db 403 ----SYVSLGLGEQRTAKARVTLTNPQAMRPRGLVTVDFGADVEVPAVKTEAVODVN 459
 QY 325 KDTVMIVNAAGMEPREVTVAQOOGTNMIVTSGLDKG 361
 Db 460 GESVVFVAVGGFVPQPVKVGRTNGKVIIEVGLKPG 496
 RESULT 13
 NCB ALCX STANDARD; PRT; 397 AA.
 ID NCB ALCX 044585;
 AC 044585;
 DT 15-DEC-1998 (Rel. 37, Last Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE Nickel-cobalt-cadmium resistance protein nccb.
 GN NCB.
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Achromobacter.

OX NCB_TaxID=515;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=31A;
 RX MEDLINE=95050278; PubMed=7961470;
 RA Schmidt T., Schlegel H.G.;
 RT "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
 RL of Alcaligenes xylosoxydans 31A.";
 CC J. Bacteriol. 176:7045-7054(1994).
 CC -1- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
 CC RESISTANCE TO NICKEL, COBALT AND CADMIUM.
 CC -1- SIMILARITY: HIGH, TO A. EUTROPHUS CNRB AND SOME, TO ALCALIGENES
 CC CZCB.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L31363; AAA65105.1; -
 CC PIR: I39579; I39579.
 CC InterPro: IPR006143; HLYD.
 CC InterPro: IPR005695; Membrane_fus2.
 CC Pfam: PF00529; HLYD: 1.
 CC TIGRfams: TIGR00999; 8a0102; 1.
 CC Plasmid; Nickel; Cobalt; Cadmium; Transport.
 CC KW SEQUENCE 397 AA; 39933 MW; C6DE57CE31499688 CRC64;
 SO
 Query Match 10.8%; Score 217.5; DB 1; Length 397;
 Best Local Similarity 28.1%; Pred. No. 4.4e-06;
 Matches 113; Conservative 51; Mismatches 149; Indels 89; Gaps 20;
 QY 12 AALAAVALVLTSSCGKGDAAQ--GAPARE-----APPVGVTVHPQYVALTYELPGR 66
 Db 16 AGVAAAAALV-----GFGARGLGSPSGAEVSKLAAPEKAAASA--PALEPAEVRIFG- 67
 QY 67 LESRTADVRAQ--VGSI-----IQRFLFGSGSY 93
 Db 68 -EYLAANINAVEPVSAGVGSVTLAPASVAAPGSEVIVISRAAGAVLRIRQKI--GDA 123
 QY 94 VRAGOPLYQIDSSYEANLESARQATLTAQATLAKADADLARYKPLVAEAVSROEYDAA 153
 Db 124 VRADVAVLAVDSP--EAAAMAAERKVAQARADLARKYE--RESSLQOQGVTPQENHESA 179
 QY 154 VTAKSAEAGVKAQAQAIKISAGINLN--RS--RITAPISGFIGOSKVSSEGLTLNAGDTVLA 211
 Db 180 RIALDVQAQAEVQRAATVAQAQAKVSSDGRSVAVVSPIGRITAOVSVLG-----A 228
 QY 212 TIRQTNMIVNTVQSAEYWKLRQIAEGKLAD--GVIAVGIKFDDGTIV-----PEKG 265
 Db 229 YVAQAELEFVAVSGAIVQV-----EAYVTADTSRIAG--SPATVILANGAPLAG 277
 QY 266 RLLEFADPVNVESTGQITLRAAVPNDON--ILMPGLVYVRILMDQVAVDNAFVVPQQAIVR 322
 Db 278 RVQAVITFTVGSARAAIV--VTPVDANSGLIVSGVQVR--LHKRAADANMSPEDAVON 335
 QY 323 GAKDTVMIVNAAGMEPREVTVAQOOGTNMIVTSGLDGDKV 364
 Db 336 LDGRDVVFVFTQCGFRPKSVLVGSRSGVAAQIISGVAPGEOV 377
 RESULT 14
 CNRB ALCEU STANDARD; PRT; 395 AA.
 ID CNRB ALCEU P37973;
 AC P37973;
 DT 01-OCT-1994 (Rel. 30, Last Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Nickel and cobalt resistance protein cnrb.
 GN CNRB.

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE005269; AAG55260.1; ALT_INIT.
 DR EMBL; AP002553; BAB34387.1; ALT_INIT.
 DR EMBL; AE015112; AAN42471.1; ALT_INIT.
 DR InterPro; IPR006143; H1YD.
 DR Pfam; PF00529; H1YD; 1.
 DR Transport; Inner membrane; Signal; Antibiotic resistance;
 KM Complete proteome.
 FT SIGNAL 1
 FT CHAIN 32 371 MACROLIDE-SPECIFIC EFFLUX PROTEIN MACA.
 SQ SEQUENCE 371 AA; 40639 MW; 25F7D3CB1A2D080F CRC64;
 Query Match 13.2%; Score 267.5; DB 1; Length 371;
 Best Local Similarity 28.1%; Pred. No. 4.6e-09;
 Matches 96; Conservative 53; Mismatches 166; Indels 27; Gaps 7;
 QY 42 APAPVGVVTVHPQTVALTELPERLESIRADYRAQVGGIIQKRLFOEGSYVRAGOPLY 101
 DQ 31 APVTVYOTLIYRPDLDQSVLATGKLDARKVDYGAQVSGQKLSVAIGDKYKDDLLG 90
 QY 102 QIDS-----STYEANLESARQAQATQAATLAKADADLARYKPLVAEAVSRQETDAV 154
 DQ 91 VIDEQAENQIKVEEATIMELRAQROQAELKLARVYSHQRLAQTOAVSLQDLDPTAA 150
 QY 155 T-AKRAEAG-----VKAQAQAIKSGACININRSRITRAPISGFIGSKVBSG-TLLNAGD 206
 DQ 151 TEMVKAQOIGTIDAQIKRNQASLDYAKTNLDYRIYAPMAGEVTOITTLLOGGVIAAQ 210
 QY 207 TVLATIRQTNPMYVNTQSAEYMKLRQIAEGKLLAADGVIAVGIKFDGTVYPEKGR 266
 DQ 211 APNLTITLADMTMLVKAQVSEADYHILKPGQAKFTYIGDPLTRYEGIKDVLPTPEK-- 268
 QY 267 LLEFADPVNVESTGOITLRAAVPNDONILMPGLYYRVLMDOVAVDNAFVPOQAVTRGAKD 326
 DQ 269 -----VNDALFYFA-RFEVPNPNGLRLMDTAQVNHQLTFDVGNVLTILPLSALGDVPD 320
 QY 327 TWMTLVN-AQSGMEPREYVAQOOGTNWIVTSGIKDGKVV 366
 DQ 321 NRYVKLLIRNGETRETEREVTIGARNDTVEIVKGLGEAGDEVVI 362
 RESULT 11
 CZCB ALCEU STANDARD; PRT; 520 AA.
 ID CZCB ALCEU
 AC P13510;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cobalt-zinc-cadmium resistance protein czcb (Cation efflux system
 DE protein czcb).
 GN CZCB.
 OS Alkaligenes eutrophus (Ralstonia eutropha).
 OC Plasmid pMO130.
 CC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 CC Burkholderiaceae; Ralstonia.
 CX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH34;
 RX MEDLINE=90017477; PubMed=2678100;
 RA Nies D.H., Nies A., Chu L., Silver S.;
 RT "Expression and nucleotide sequence of a plasmid-determined divalent
 RT cation efflux system from Alkaligenes eutrophus";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7351-7355(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH34;
 RA van der Leije D., Schwuchow T., Wuerztz S., Schwidetzky U.,
 RA Baeyens W., Scheel P.O., Nies D.H.;
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.

CC -I- FUNCTION: CZCA AND CZCB TOGETHER WOULD ACT IN ZN(2+) EFFLUX
 CC NEARLY AS EFFECTIVELY AS THE COMPLETE CZC EFFLUX SYSTEM (CZCABC).
 CC THE CZCB PROTEIN IS THOUGHT TO FUNNEL ZN(2+) CATIONS TO THE CZCA
 CC TRANSPORT PROTEIN.
 CC -I- INDUCTION: BY CADMIUM, COPPER AND ZINC.
 CC -I- SIMILARITY: SOME, TO A. EUTROPHUS CNRB.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X98451; CA67083.1; -.
 DR PIR; B33830; B33830.
 DR InterPro; IPR006143; H1YD.
 DR InterPro; IPR005695; Membrane_fus2.
 DR Pfam; PF00529; H1YD; 1.
 DR TIGRPFams; TIGR00999; 8a0102; 1.
 KM Plasmid; Transport; Zinc; Cobalt; Cadmium resistance.
 SQ SEQUENCE 520 AA; 54587 MW; FC3D347F87A7A79 CRC64;
 Query Match 11.9%; Score 241; DB 1; Length 520;
 Best Local Similarity 25.9%; Pred. No. 2.4e-07;
 Matches 103; Conservative 56; Mismatches 168; Indels 70; Gaps 12;
 QY 9 MRPAALAAVALVLSGCKGDAAGGQPA-----GREAPAV----- 46
 DQ 126 VRATGESQALKPFVVS-----GDALIESQOPVAEPHVPDVTANVTLPGSSSPLAVRLSKEG 180
 QY 47 -----VGVV--TVHPQTVALTELPG--PLESLRTADYRAQVGGIIQKRLFOEG 91
 DQ 181 KIETLADQAKTGGVTVQTAGSANVQAGVQFPEIRFENEDKTAHVYPLAGVBSVPANIG 240
 QY 92 SYRAGOPLYQIDSTYEANLESARQAQATQAATLAKADADLARYKPLVAEAVSRQETD 151
 DQ 241 QYKKGQVLAIVAST-----QLSDQRSSELLAAQRLDLARTYREKLMQOKISABEDYL 296
 QY 152 AAVTAKRSAGYKAAQAIKSGIN-----INRSRTADISGFIGSKVSEGLTLNAGD 206
 DQ 297 SARMLQEAQISVONQOKLTALGASNSPALNRYELRAPFAGMIVKHSLSGAIV--AD 354
 QY 207 TVLATIRQTNPMYVNTQSAEYMKLRQIAEGKLL--AADGVIAVGIKFDGTVYPEK 264
 DQ 355 NAVVFTLSDISSVWAEFVVSADKOVERR--IGEKASINSASDVKA-----DGTV---- 402
 QY 265 GRLLFADPVNVESTGOITLRAAVPNDONILMPGLYYRVLMDOVAVDNAFVPOQAVTRGAK 324
 DQ 403 ---SYVSSLGEGQTRTAKAVTLTNPMAMRPGLFVYVDVFGADVVPVAVKTEAVQDVN 459
 QY 325 KDTVMIVNAQSGMEPREYVAQOOGTNWIVTSGIKDG 361
 DQ 460 GESVVFVAQVGFVQPVKVRGRTNGKVIIEVGLKPG 496
 RESULT 12
 CZCB ALCEU STANDARD; PRT; 520 AA.
 ID CZCB ALCEU
 AC P94176;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cation efflux system protein czcb.
 GN CZCB.
 OS Alkaligenes sp. (strain CT14).
 CC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 CC Alcaligenaceae; Alkaligenes.
 CX NCBI_TaxID=512;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96219090; PubMed=8829543;
 RX

OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [2]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=K12 / W3104;
 RX MEDLINE=21429237; PubMed=11544226;
 RA Kobayashi N., Nishino K., Yamaguchi A.;
 RT "Novel macrolide-specific ABC-type efflux transporter in Escherichia coli";
 RL J. Bacteriol. 183:5639-5644(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1234-1238(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajinaka M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:137-155(1996).
 CC -1- FUNCTION: Efflux transporter for macrolide antibiotics.
 CC -1- SUBUNIT: Interacts with macb.
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB071145; BAB64541.1; ALT_INIT.
 DR EMBL; AB000189; AAC73965.1; ALT_INIT.
 DR EMBL; D90725; BAA35597.1; -.
 DR EcoGene; EGI3694; maca.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KM Transport: Inner membrane; Signal; Antibiotic resistance;
 KW Complete proteome.
 FT SIGNAL 1 31
 FT CHAIN 32 371 MACROLIDE-SPECIFIC EFFLUX PROTEIN MACA.
 SQ SEQUENCE 371 AA; 40624 MW; 8BPF287DA03B92AF9 CRC64;
 Query Match 13.3%; Score 269.5; DB 1; Length 371;
 Basic Local Similarity 28.1%; Pred. No. 3.5e-09;
 Matches 96; Conservative 54; Mismatches 165; Indels 27; Gaps 7;
 QY 42 APAPVGVVTVHPOTVALVELPGLRLSLRTADYRAVGIIQRLRFGSGSYVRAQGPXY 101
 DB 31 APVPTVTLVRPDDLOQSVLATKDLARKVDVGAQVSGQKTLVLAIGDKYKKDQLLG 90
 QY 102 QIDS-----STYEANLESRAQATLAQATLAKADADLARYKELVAEAVSRQEDYAAV 154
 DB 91 VIDEQAENQIKVEATLMELEAQRQAELKLAARYTSQGLAQTAKVSGQDDLTAA 150
 QY 155 T--AKRAEAG-----VKAQAALAKSGINLNSRTAPISGPIGSGVSEB--TLNAGD 206
 DB 151 TEAAVVKOAQIGTIDAOIKRQASLDATKNTLDYTRIIVAPMAGEVQTITTLQGGIVIAAQ 210

QY 207 TVVLTATIRQTNPMYVNTVQASAEVWKLRRQIAEGKLLAADGVIAWGKFDGTVYPEGR 266
 DB 211 APNLTITLADMSAMLVKQVSEADVIHLKPGQKAMFTVLGDPFLRREGQIKDVLPTPERK-- 268
 QY 267 LFPADPVNNSGQITILRAVPNDQNLMPGLVYRVLMDOVAANDNAVVPQQAATRKAD 326
 DB 269 -----VNDAIFFYA--RFEVPPNPNGLRLDMTAQVHIQLTDVKNVLTITPLSALGDVPGD 320
 QY 327 TVMIVN--AAGGMEPREVTAQOQGTWMTVTSGLKDGDKVTV 366
 DB 321 NRYVVKLLRNGETREREVTTIGANDIDVEIVKGLRAGDEVVI 362
 RESULT 10
 MACA_EC057
 ID MACA_EC057 STANDARD; PRT; 371 AA.
 AC P68410;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Macrolide-specific efflux protein maca precursor.
 GN MACA OR Z1115 OR ECS0964 OR SF0838.
 OS Escherichia coli O157:H7, and
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334, 623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grodbeck E.J., Davis N.W., Lim A., Dimaranta E.T., Potamouis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Shigella flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=2272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Ou D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 through comparison with genomes of Escherichia coli K12 and O157";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 CC -1- FUNCTION: Efflux transporter for macrolide antibiotics (By
 similarity).
 CC -1- SUBUNIT: Interacts with macb (By similarity).
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ414148; CAC90193.1; -
 CC EMBL: AE013884; AAM86365.1; -
 CC PIR: AF0166; AF0166.
 CC InterPro: IPR006143; H1YD.
 CC Pfam: PF00529; H1YD; 1.
 CC Transport: inner membrane; Signal; Antibiotic resistance;
 CC Complete proteome.
 CC SIGNAL 1 26 POTENTIAL.
 CC CHAIN 27 371 PROBABLE MACROLIDE-SPECIFIC EFFLUX
 CC PROTEIN MACA.
 CC
 CC SEQUENCE 371 AA; 40186 MW; 13173F46A8696E3E CRC64;
 SO
 Query Match 14.7%; Score 297; DB 1; Length 371;
 Best Local Similarity 29.8%; Pred. No. 8.4e-11;
 Matches 112; Conservative 63; Mismatches 151; Indels 50; Gaps 12;
 QY 16 AANVALVSSCGKGDAAOGQAPAGREAPVGVVTHPQVAL-----TVELPRL 68
 DB 13 AAIIVLTL-----GGFFIARHLMAP-----VPVNVYVKKVHRDLQONVALATGKLD 58
 QY SLRFADVAQVGGIIQKRLFOEGSYVRAGOPLYQIDS-----STEANLESARAO LAT 121
 DB 59 AVRKVDVAQVSGLEKLYVEIGHVKGGLAMIDPQQAQNIKEVETATIQDLMAQRIO 118
 QY 122 AQAATLAKADADLARYKPLVAEAVSRQEDYDAVT--AKRSAEAG-----VKAQAQAIKSA 174
 DB 119 AKAEHLHATVTLGRQONLAKIQVNSRQDLQAVTDLAVKNAKVGITIDQINKAKASLDTA 178
 QY 175 GINLNRSITAPISGFIQSKVSEB-TLLNAGDTIVLATIQTNPMYNNVTSQASSEVWKL 233
 DB 179 KINLDYTOISAPMGDDVQITTLQOQTVIAAQAPNITLLADMSTMLVQAQVSEADVTNL 238
 QY 234 RROI-AEGLKLAADGVIAVGIKPDGTVYPEKGRLLFPADPVNVESTQGITLRAVPNDQN 292
 DB 239 KPGKASTTVIGDEKRRSGV-LKDILPTPK-----VNDAI-FYSARFEVNPDR 287
 QY 293 ILMGLVYRVLMDQVAVDNAAFVPOQAV--TRGAKDTVMIVNAQGMREPRETVAAQOQT 350
 DB 288 LLRLQMTAQVSIQLANVDQAVVILPLAALGDELGSNRQVTVLKGKKEKEKREVTIGIRNV 347
 QY 351 NWIVTSGIKDGDKVIV 366
 DB 348 DAOVISGLSVEGDVIV 363
 RESULT 8
 ID 1894_HAEIN STANDARD; PRT; 382 AA.
 AC Q57500;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein H10894.
 GN H10894.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 CX NCI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
 RA Utecherack T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32771; AAC22554.1; -
 CC TIGR: H10894; -
 CC Pfam: PF00529; H1YD; 1.
 CC Hypothetical protein; Transmembrane; Complete proteome.
 CC TRANSMEM 1 21 POTENTIAL.
 CC SEQUENCE 382 AA; 41409 MW; 7974AA9D53F0BB5E CRC64;
 SO
 Query Match 14.2%; Score 286; DB 1; Length 382;
 Best Local Similarity 25.2%; Pred. No. 3.9e-10;
 Matches 105; Conservative 66; Mismatches 148; Indels 98; Gaps 14;
 QY 6 PKMRAALAAVALVUSSCGKGDAAGQAPAGREAPVGVVTHPQVALTVELPG 65
 DB 18 FNMIRGVNISRALA-----GMP--ESSSPVT-ALEVQPREMTVINTTG 58
 QY 66 RLSELRADVAQVGGIIQKRLFOEGSYVRAGOPLYQIDSSTEANLESARAO LAT 125
 DB 59 LVRRNQAMSLSTQAGAVSQVLVONGQNVKKGEVLVLDSSVEGANIQAQAQSLAROT 118
 QY 126 LAKADADLARYKPLVAEAVSRQEDYDAVTAKRSAGKVAQAQAIKSAGINLRSRTA 185
 DB 119 -YQRYVGLINSAVSRQEDNKAAYDAQVASISLSKAIE-----RRRTVA 164
 QY 186 PISGFIQSKVSEBETLNAAGDTIVLATIQTNPMYNNVTSQASSEVWKL 245
 DB 165 PFDDKAAIVKINVGQYVNVG--TEIVREDTISSMKVDFALSQNDLKLH---IGQRTVA 218
 QY 246 DGVIAVGIKPDGTVYPEKGRLLFPADPVNVESTQGITLRAAV--PNQNIILMGLVYRVL 304
 DB 219 TTDARLGEFT-----SARITAIERAINSTGLVAVQATFPEDEDHKLKSGFSLRI 270
 QY 305 DOVAVDNAFVVPQAVT-----RGAKDTVMIVN 332
 DB 271 ALPTETNQVVVPPQVVAISYNNYGEIAYLLPELSEEEKMGSGNEKLDRLYRAKQITVFTKD 330
 QY 333 AOGMEREPTVAAQOQGTNMIIVSGLDGDKVVEGISAGITCAKVTTPREMASS 389
 DB 331 RQG-----VYQLOQNE-----VYVGDKIITGQO--GIGNSLV--EWIKD 369
 RESULT 9
 ID MACA_ECOLI STANDARD; PRT; 371 AA.
 AC P75830;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Macrolide-specific efflux protein maca precursor.
 GN MACA OR B0878.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

QY 205 GDTVLTATIRCTNPMYVNTOSASEVMKLRQIAEGKLLAADGVIAVGKPDGTGYPEK 264
 Db 194 NOADSLVTVQRLDPIYVLTOSVDPFLRKEEVAISGQIKQVGSPTPLMNGREYSQT 253
 QY 265 GRLLFADVNESTQITLRAAVPNDONILMPGLVYRVLMDOVAVDNAFVVPQOAVTRGA 324
 Db 254 GTLKESDPTVDETTSSVTLRAIFRPNGDILPQMVTYALVBSGRONVLVPQEGVTNNA 313
 QY 325 --KDTVMIVNAOGMEPREVTVAOQGGTNMIVTSGLKDGKRVVVGIS--IAGITGAKKV 380
 Db 314 QGKATATILDKDDVQVQLREIEASKAIGDQWVVTSGIQGDRVIVSGLQIRIRGIARAIS 373
 QY 381 TPKEWASSENQ 391
 Db 374 SSGENASTESK 384

RESULT 6

YEGM_ECOLI STANDARD; PRT; 415 AA.
 ID YEGM_ECOLI
 AC P76397;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yegm precursor.
 GN YEGM OR B2074.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayaishi K., Inada T., Isono K.,
 RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
 RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
 RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
 RA Sivasubramanian S., Tagami H., Takeda J., Takemoto K., Wada C.,
 RA Yamamoto Y., Horikuchi T.,
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map."
 RL DNA Res. 3:379-392(1996).
 CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

DR EMBL: AE000297; AAC75135.1; ALT INIT.
 DR EMBL: D90845; BAA15928.1; ALT_INIT.
 DR Ecogene; EG14056; yegm.
 DR InterPro; IPR006143; HLYD.
 DR Pfam; PF00529; HLYD; 1.
 KM Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 415 HYPOTHETICAL PROTEIN YEGM.
 FT SEQUENCE 415 AA; 44464 MW; 26240DBBDEBIASF CRC64;

Query Match 21.4%; Score 432; DB 1; Length 415;
 Best Local Similarity 32.2%; Pred. No. 1,1e-18;
 Matches 119; Conservative 63; Mismatches 168; Indels 20; Gaps 6;

QY 34 GGPPAGRAPAPVGVVTVHPQVALTVELPGRLESRTADVPAQVGSIIQRLPQEGSY 93
 Db 49 GGRGMSGRLAPVQATVAVQAVPRVITGIGTTAANTVTVRSVDQLALHFQEQO 108
 QY 94 VRAQOPLYQIDISTYEANLESAPAQLATAQATLAKADADLARVPLVAEAVSRQEDYAA 153
 Db 109 VKAGDPLAEIDPQGFVKYALAAQAOQLAKDKXTLANARDLARVQQLAKTNLVSGELDAQ 168
 QY 154 VTAKRSAPGVKAQAQAIAKSGINLSRTAPISGFIQGSVSEGTILNAGDTTVLATI 213
 Db 169 QALVSETEGITKADAEASVAAQQLDWSRITAPADGVGVLKQVDVGNQISSGDTTGIVVI 228
 QY 214 RQTPMYVNTVQASAEVMKLRQIAEGKLLAADGVIAVGI-KPFDGTGYPEKGLLPADP 272
 Db 229 TQTHPIDLVFTLPESDIAVTVQAOAKGKPLVEAMDRTNSKKLSEGT-----LLSIDN 281
 QY 273 VVNESTQITLRAAVPNDONILMPGLVY--RVLMDOVAVDNAFVVPQOAVTRGAK-DTVM 329
 Db 282 QIDATTTGIIKAKAFNNQDDALFPNQVYNAMLVD--TEQNAVVIPTALQMGNGHFVW 339
 QY 330 IVNAOGMEPREVTVAOQGGTNMIVTSGLKDGKRVVVGI-----SIAGITGAKKVT 382
 Db 340 VLNSENKVSRLVPGIQDSQKVIRAGISAGDEVTVDGIDRLTEGAKVEVEAQSATTP 399
 QY 383 KEWASSENQ 392
 Db 400 EKAATREYA 409

RESULT 7

MACA_YERPE STANDARD; PRT; 371 AA.
 ID MACA_YERPE
 AC P56411;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable macrolide-specific efflux protein maca precursor.
 GN MACA OR YF01364 OR Y2814.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.

OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Fellwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

RP STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Bourin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Herscherson J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Miles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.,
 RT "Genome sequence of Yersinia pestis KIM."
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -1- FUNCTION: Efflux transporter for macrolide antibiotics (By similarity).
 CC -1- SUBUNIT: Interacts with macB (By similarity).


```

RT Escherichia coli.":
RL Mol. Gen. Genet. 230:230-240(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-96 FROM N.A.
RC STRAIN=K12;
RA Klein J.R., Henrich B., Plapp R.;
RT "Molecular cloning of the envC gene of Escherichia coli.";
RL Curr. Microbiol. 21:341-347(1990).
CC -1- FUNCTION: MAY AFFECT SPECIFIC MEMBRANE FUNCTIONS, SUCH AS SEPTUM
CC FORMATION DURING CELL DIVISION, AND CELL MEMBRANE PERMEABILITY.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (potential).
CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96848; AAA02931.1; -
DR EMBL; X57948; CAA41016.1; -
DR EMBL; U18997; AAB58069.1; -
DR EMBL; AE000405; AAC76297.1; -
DR PIR; C65119; C65119.
DR EcoGene; EG10266; acrc.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Cell division; Signal; Inner membrane; Lipoprotein; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 1 23
FT LIPID 24 385
FT CONFLICT 191 212
FT CONFLICT 313 314
FT SEQUENCE 385 AA; 41317 MW; EDBF82C83CF42A9 CRC64;
SQ
Query Match 37.2%; Score 750.5; DB 1; Length 385;
Best Local Similarity 42.1%; Pred. No. 1.7e-37;
Matches 168; Conservative 74; Mismatches 122; Indels 35; Gaps 6;
OY 17 AVALVLSCGKGDAAGGPGAREAPVGVTVH-PQTVALTV-ELPGLSLRTA 73
DB 17 SAALLAGCNDGGEBAHGEPO-----VTVHVKTAPLEVKELGRNNAVYIA 65
OY 74 DVRAQVGIIQRLFOGSGYVPAQOPLYOIDSSTYEANLESARAQALTAQATLAKADL 133
DB 74 DVRAQVGIIQRLFOGSGYVPAQOPLYOIDSSTYEANLESARAQALTAQATLAKADL 133
OY 66 EVRPVSGIVINRNFTSGSDVQAGSLYQIDPATQANYDSAKGELASAAAIAHLTV 125
DB 66 EVRPVSGIVINRNFTSGSDVQAGSLYQIDPATQANYDSAKGELASAAAIAHLTV 125
OY 134 ARYPVVAEAVSROEYVAATVAKRSAGYKAAQALIKSAGININRSRTAPISGFIQ 193
DB 134 ARYPVVAEAVSROEYVAATVAKRSAGYKAAQALIKSAGININRSRTAPISGFIQ 193
OY 126 KRIYPLVGTXTYISQEDQALADARQADAAVIAAATVESAIRINAIYKVTAPISGRIGK 185
DB 126 KRIYPLVGTXTYISQEDQALADARQADAAVIAAATVESAIRINAIYKVTAPISGRIGK 185
OY 194 SKVEGTLNAGDTTVLATITROTNPVYVNTOSASEWKLROIAEGKILAADGVIAVG 253
DB 194 SKVEGTLNAGDTTVLATITROTNPVYVNTOSASEWKLROIAEGKILAADGVIAVG 253
OY 186 STVEGALVTNGQTTTELATVQQLDPIYDVTOSSDFRLKQSVQGLHKNMSTNVEL 245
DB 186 STVEGALVTNGQTTTELATVQQLDPIYDVTOSSDFRLKQSVQGLHKNMSTNVEL 245
OY 254 KFDGCTVPEKGRLLFADPVVNESTGQITTLRAAVNDONILMPLGYVAVLMDQVAVDNF 313
DB 254 KFDGCTVPEKGRLLFADPVVNESTGQITTLRAAVNDONILMPLGYVAVLMDQVAVDNF 313
OY 246 VMENGQTYPLKQTLQFSDVTVDSTGSLTLRAVFPNPQHTLLPGMFVARIDEGVQDAI 305
DB 246 VMENGQTYPLKQTLQFSDVTVDSTGSLTLRAVFPNPQHTLLPGMFVARIDEGVQDAI 305

```

```

OY 314 VVPOQAVTRGAKD--TVMIVNAGGMEPREVTVAOQOGTWNIVTSGLKDGKVVVEGISI 371
DB 306 LIPOQVSRTRPGATVTLVNDKSGQVEARPVVASQALGDKWLISEGKSGDQIVVSGL-- 363
OY 372 AGITGAKKVTPEKEMASSENQAAAPQSGVQTAASEKTKASE 410
DB 364 -----OKARPG-----QVATVTDTPADTASK 385

RESULT 5
YHIU ECOI STANDARD; PRT; 385 AA.
ID YHIU ECOI
AC P37636;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein yhiu precursor.
GN YHIU OR B3513.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (potential).
CC -1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00039; AAB18489.1; -
DR EMBL; AE000427; AAC76538.1; -
DR PIR; S47733; S47733.
DR EcoGene; EG12240; yhiU.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Signal; Inner membrane; Lipoprotein;
KW Complete proteome.
FT SIGNAL 1 20
FT CHAIN 1 20
FT LIPID 21 385
FT LIPID 21 21
FT SEQUENCE 385 AA; 41190 MW; 2C825B6CDE15C70F CRC64;
SQ
Query Match 32.5%; Score 656; DB 1; Length 385;
Best Local Similarity 38.5%; Pred. No. 6.4e-32;
Matches 143; Conservative 71; Mismatches 153; Indels 4; Gaps 2;
OY 25 CGKGGDAAGGPGAREAPVGVTVH-PQTVALTV-ELPGLSLRTADVRAQVGIIQ 84
DB 14 CGAMLTACDDKSAMNAAAMPYGVTVLSFGSVNVVSELGRITVPYVAELRPVGGII 73
OY 85 KRLFOGSGYVPAQOPLYOIDSSTYEANLESARAQALTAQATLAKADLARYKPLVAEEA 144
DB 85 KRLFEEDKYNQGGSLYQIDPAPLQAEINSAKSLALSTASVARTFNRQASLLKTNV 133
OY 145 VSRQEVYVAATVAKRSAGYKAAQALIKSAGININRSRTAPISGFIQSGSVSSGTLNA 204
DB 134 VSRQDYTAARTQNLNEAEANVTAAVAEQAATINLOVANVTSPITGVSGKSSVTGALVTA 193

```


Dd	61	VELPGRLESLRTNADVRAQVGIIQKRLPQEGSVVRAGQPIYQJDSSTVEAGLESAARQLA	120
Qy	121	TAQATLAKADADLARYKPLVAAEVASRQEDYDAVTAKRSAAGVKAQAAIKAGINLNR	180
Dd	121	TAQATLAKADADLARYKPLVASDAIISKQEDYDAVTAKRSAEASVKAQAQAIKAGINLNR	180
Qy	181	SRITAPISGFIQSQSKVSEGGTLNAGDTYVLAIRQNPMPVYVNTQSASEPMKRRQIAEG	240
Dd	181	SRITAPISGFIQSQSKVSEGGTLNAGDTYVLAIRQNPMPVYVNTQSASEPMKRRQIAEG	240
Qy	241	KLLAADGVTAAGIKFPDGTIVYPEKGLLFPADPVNVESTGQITLRAVPNDQNLMPGLYV	300
Dd	241	KLLAADGAIVAGIKFPDGTIVYPEKGLLFPADPVNDSTGQITLRAAVSNDQNLMPGLYV	300
Qy	301	RYLMDQVAVDNAFVVPQQAVTGAKDQTVMIVNAQGMPEPREVTVAQOQGNTMIVTSGLKD	360
Dd	301	RYLMDQVAVDNAFVVPQQAVTGAKDQTVMIVNAQGMPEPREVTVAQOQGNTMIVTSGLKD	360
Qy	361	GDQYVVEGISTINGTCAKKTVPTEKMASSNNQAAAPQSGVQTASAEATAGTAEE	412
Dd	361	GDQYVVEGISTAGMTGAKKVTPEKMASSNNQAAAPQAGVQTASEAKPASEAK	412
RESULT 2			
ACRA	ECOLI	STANDARD;	PRT; 397 AA.
ID	ACRA_ECOLI		
AC	P31223;		
DT	01-JUL-1993 (Rel. 26, Created)		
DT	01-JUL-1993 (Rel. 26, Last sequence update)		
DT	01-FEB-2003 (Rel. 41, Last annotation update)		
DE	Acriflavine resistance protein A precursor.		
GN	ACRA OR MTCA OR LIR OR B0463 OR Z0578 OR EC05016.		
OS	Escherichia coli, and		
OS	Escherichia coli, and		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Escherichia.		
OX	NCBI_Taxid=562, 83334;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12;		
RA	Xu J., Bertrand K.P;		
RT	"Nucleotide sequence of the acraB operon from Escherichia coli.";		
RL	Submitted (May-1993) to the EMBL/Genbank/DBD databases.		
RN	(2)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / M4573;		
RC	MEDLINE=944012493; PubMed=8407802;		
RA	Ma D., Cook D.N., Alberti M., Pon N.G., Nikaido H., Hearst J.E.;		
RT	"Molecular cloning and characterization of acra and acrc genes of		
RT	Escherichia coli.";		
RL	J. Bacteriol. 175:6299-6313(1993).		
RN	(3)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MG1655;		
RC	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.";		
RL	Science 277:1453-1474(1997).		
RN	(4)		
RP	SEQUENCE FROM N.A.		
RP	Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,		
RA	Duncan M., Federpiel N., Hyman R., Kalman S., Komp C., Kurd O.,		
RA	Lew H., Lin D., Natsuch A., Oelfer P., Schiann S., Davis R.W.;		
RL	Submitted (Jan-1997) to the EMBL/Genbank/DBD databases.		
RN	(5)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / EDL93 / ATCC 700927;		
RC	MEDLINE=21074935; PubMed=11206551;		
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,		
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,		

RA Postel Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalina E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.,
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [6]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11256796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Yamaguchi M., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,
RA Kunita S., Shiba T., Hattori M., Shinagawa H.,
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL Nature Res. 8:11-22(2001).
RN [7]
RN CHARACTERIZATION.
RX MEDLINE=95379493; PubMed=7651136;
RA Ma D., Cook D.N., Alberti M., Pon N.G., Nikaio H., Hearst J.E.,
RT "Genes acra and acrb encode a stress-induced efflux system of
RT Escherichia coli.";
RL Mol. Microbiol. 16:45-55(1995).
RN [8]
RN PROCESSING, AND INTERACTION WITH ACRB.
RX MEDLINE=20381028; PubMed=10920254;
RA Kawabe T., Fujihira E., Yamaguchi A.,
RT "Molecular construction of a multidrug exporter system, ACrAB:
RT molecular interaction between AcraA and Acrb, and cleavage of the
RT N-terminal signal sequence of AcraA.";
RL J. Biochem. 128:195-200(2000).
CC -I- FUNCTION: AcrAB is a drug efflux protein with a broad substrate
CC specificity.
CC -I- SUBUNIT: Interacts with acrb.
CC -I- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor.
CC -I- SIMILARITY: BELONGS TO THE ACRA/ACRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M94248; AAA23410.1; -;
DR EMBL, U00734; AAA67134.1; -;
DR EMBL, AE000152; AAC73565.1; -;
DR EMBL, U82664; AAB40217.1; -;
DR EMBL, AE005225; AAG54812.1; -;
DR EMBL, AP002551; BAB33939.1; -;
DR PIR, A36938; A36938.
DR PIR, D90693; D90693.
DR PIR, H85543; H85543.
DR Ecogene; EG11703; acra.
DR InterPro; IPR006143; HlyD.
DR Pfam; PF00529; HlyD_1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Transport; Inner membrane; Signal; Antibiotic resistance;
KW Lipoprotein; Complete proteome.
FT SIGNAL 1 24
FT CHAIN 25 397
FT LIPID 25 25 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 397 AA; 42196 MW; 5B81DD5BC2B0A077 CRC64;
Query Match 37.8%; Score 762.5; DB 1; Length 397;
Best Local Similarity 42.2%; Pred. No. 3.5e-38;
Matches 168; Conservative 76; Mismatches 129; Indels 25; Gaps 6;
11 AALAAVALVLSGKGGDAGGQDPAAGRBAPAVGVTVTHQTVALLVLPRLSL 70

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2003, 13:56:01 ; Search time 17 Seconds
(without alignments)
1139.706 Million cell updates/sec

Title: US-09-889-756A-2

Perfect score: 2019
Sequence: 1 MAFYAFAMRAALAAVAL.....AAPQGVQTAASEATSEAE 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1938	96.0	412	1 MTRC_NEIGO	P43505 neisseria g
2	762.5	37.8	397	1 ACRA_ECOLI	P31223 escherichia
3	753	37.3	383	1 MEXA_PSEAE	P54277 pseudomonas
4	750.5	37.2	385	1 ACRA_ECOLI	P24180 escherichia
5	656	32.5	385	1 YHUI_ECOLI	P37636 escherichia
6	432	21.4	415	1 YEGM_ECOLI	P76397 escherichia
7	297	14.7	371	1 MACA_YERPE	P58411 yersinia pe
8	286	14.2	382	1 H894_HAEIN	Q57500 haemophilus
9	269.5	13.3	371	1 MACA_ECOLI	P75830 escherichia
10	267.5	13.2	371	1 MACA_ECOLI	P58410 escherichia
11	241	11.9	520	1 CZCB_ALCEU	P33510 alcaligenes
12	236	11.7	520	1 CZCB_ALCEU	P94176 alcaligenes
13	217.5	10.8	397	1 NCCE_ALCEX	Q44585 alcaligenes
14	194	9.6	395	1 CNRB_ALCEU	P37973 alcaligenes
15	187	9.3	430	1 SLUB_SALTY	Q92480 salmonella
16	184	9.1	367	1 NOLF_RHIME	P25196 rhizobium m
17	183	9.1	328	1 Y873_YERPE	Q93447 yersinia pe
18	178.5	8.8	343	1 YGCR_ECOLI	P32716 escherichia
19	173.5	8.6	332	1 YBHG_ECOLI	Q87496 escherichia
20	172.5	8.5	332	1 YBHG_ECOLI	Q87496 escherichia
21	171.5	8.5	332	1 YBHG_ECOLI	Q87496 escherichia
22	169	8.4	285	1 YDHJ_ECOLI	P75777 escherichia
23	168.5	8.3	331	1 YBHG_ECOLI	P76185 escherichia
24	168.5	8.3	331	1 YBHG_ECOLI	Q82879 salmonella
25	162	8.0	407	1 CUSB_ECOLI	Q82879 salmonella
26	156.5	7.8	390	1 EMRA_HAEIN	P77233 escherichia
27	154.5	7.7	310	1 YHCO_ECOLI	P49428 haemophilus
28	150.5	7.5	390	1 EMRA_ECOLI	P76303 escherichia
29	142.5	7.1	378	1 YIBH_ECOLI	P32107 escherichia
30	138.5	6.9	355	1 YHII_ECOLI	P37626 escherichia
31	136.5	6.8	334	1 YOH3_RHIME	Q92444 rhizobium m
32	134.5	6.7	544	1 CH63_RHIME	Q93040 rhizobium m
33	133.5	6.6	387	1 EMRK_ECOLI	P52599 escherichia

34	130	6.4	2453	1 NCRI_MOUSE	Q60974 mus musculus
35	128	6.3	338	1 YBHG_THLFE	Q919d4 thlobacillu
36	127.5	6.3	329	1 AN36_HELPY	P94851 helicobacte
37	125.5	6.2	378	1 YVAV_ECOLI	P37683 escherichia
38	123	6.1	432	1 APRE_PSEAE	Q03025 pseudomonas
39	123	6.1	525	1 YAO2_RHIME	Q52969 rhizobium m
40	122.5	6.1	440	1 CYAD_BORPE	P11091 bordetella
41	121	6.0	904	1 IF2_XANAC	Q8P355 xanthomonas
42	120	5.9	545	1 CH61_RHIME	P35469 rhizobium m
43	120	5.9	545	1 CH64_RHIME	Q92294 rhizobium m
44	118.5	5.9	544	1 CH63_RHIME	Q91690 rhizobium m
45	117.5	5.8	549	1 CH65_RHILLO	Q981j9 rhizobium m

ALIGNMENTS

RESULT 1	ID	MTRC_NEIGO	STANDARD	PRT	412 AA
AC	P43505				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	Membrane fusion protein mtrc precursor.				
GN	MTRC.				
OS	Neisseria gonorrhoeae.				
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;				
OC	Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=485;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FA19;				
RX	MEDLINE=94254732; PubMed=8196548;				
RA	Pan W., Spratt B.G.;				
RY	"Regulation of the permeability of the gonococcal cell envelope by the mtr system."				
FL	Mol. Microbiol. 11:769-775(1994).				
CC	-1- FUNCTION: CELL MEMBRANE LIPOPROTEIN, INVOLVED IN CELL MEMBRANE PERMEABILITY TO HYDROPHOBIC COMPOUNDS SUCH AS ANTIBIOTICS, DYES AND DETERGENTS.				
CC	-1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid anchor (Potential).				
CC	-1- SIMILARITY: BELONGS TO THE ACRA/ACRE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between The Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL, U14993; AAA80193.1; -				
DR	InterPro: IPR006143; H1YD.				
DR	Pfam: PF00529; H1YD; 1.				
DR	PROSITE: PS00013; PROKAR LIPOPROTEIN, 1.				
KW	Signal; Inner membrane; Lipoprotein.				
FT	SIGNAL 1				
FT	POTENTIAL 24				
FT	CHAIN 25				
FT	LIPID 25				
FT	MEMBRANE FUSION PROTEIN MTRC.				
FT	N-ACYL DIGLYCERIDE (POTENTIAL).				
SO	SEQUENCE 412 AA; 42773 MW; 979FAFBCFAE321BA CRC64;				
QY	Query Match	96.0%	Score 1938;	DB 1;	Length 412;
QY	Best Local Similarity	95.4%	Pred. No. 2.2e-107;		
QY	Matches 393;	Conservative	9;	Mismatches 10;	Indels 0;
QY				Gaps 0;	
DB	1 MAFYAFAMRAALAAVALVLSGCGGDAAGGAGGEAPAPVGVVTPHQTVALT 60				
DB	1 MAFYAFAMRAALAAVALVLSGCGGDAAGGAGGEAPAPVGVVTPHQTVALT 60				
QY	61 VELPGRLESIRTDVDAOVGGIIQKRLFOEGSVYRAGOPLYQIDSSYEAANLSARAOLA 120				